# EXHIBIT 1


# Patent Business Goals (PBG) - Final Rule

Hiram Bernstein
Senior Legal Advisor
Office of Patent Legal Administration (OPLA), USPTO
(703) 305-8713 hiram.bernstein@uspto.gov

*PBG - Final Rule webpage* **http://www.uspto.gov/web/offices/dcom/olia/pbg/index.html**

**August 1, 2001**

# § 1.105 - Requirement for Information Implementation

- What is the plan to ensure proper implementation?
  - Initial Training
    - Examiners should narrowly define the scope of a requirement to minimize the burden on an applicant and because overly broad requirements
      - are difficult for applicants to comply with
      - create undue burdens on both the applicant and the examiner
      - create excessive amounts of information
    - A requirement under 37 CFR § 1.105 should not be a fishing expedition
  - **Supervisory Review**
    - **Each Technology Center will provide supervisory review and authorization over all rule § 1.105 requirements prior to mailing**
    - **Directors will control the implementation within each Technology Center**
  - Follow-Up Monitoring & Evaluation of Results
    - Effectiveness of Results
    - Reasonableness of Requirements
    - Evaluated within and across Technology Centers.

# APPENDIX A

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-000100US | SYSTEMS & METHODS FOR IMAGING CORNEAL PROFILES | John S. Shimmick<br>Stephen J. Hinkson<br>Charles R. Munnerlyn | 08/863665<br>05/27/1997 | | |
| 018158-000110US | Systems and Methods for Imaging Corneal Profiles | John S. Shimmick<br>Stephen J. Hinkson<br>Charles R. Munnerlyn | 09/083773<br>05/22/1998 | 6302876<br>10/16/2001 | |
| 018158-000120US | Systems and Methods for Imaging Corneal Profiles | John S. Shimmick<br>Stephen J. Hinkson<br>Charles R. Munnerlyn | 09/565764<br>05/05/2000 | 6520958<br>02/18/2003 | |
| 018158-000130US | Systems and Methods for Imaging Corneal Profiles | John S. Shimmick<br>Stephen J. Hinkson<br>Charles R. Munnerlyn | 09/565851<br>05/05/2000 | 6315413<br>11/13/2001 | |
| 018158-000140US | Systems and Methods for Imaging Corneal Profiles | John S. Shimmick<br>Stephen J. Hinkson<br>Charles R. Munnerlyn | 09/565774<br>05/05/2000 | 6605081<br>08/12/2003 | |
| 018158-000200US | ELECTROSTATIC PRECIPITATOR FOR A GAS DISCHARGE LASER | David I. Cullumber | 08/688715<br>07/31/1996 | 5729564<br>03/17/1998 | |
| 018158-000300US | Laser REfractive Surgery Station | Mizuno Katsuyoshi<br>Ihara Akira | 08/752662<br>11/19/1996 | 5795351<br>08/18/1998 | |
| 018158-000600US | TREPHINE FOR LAMELLAR KERATECTOMY | Stephen J. Koons | 60/077642<br>03/11/1998 | | |
| 018158-000610US | Trephine for Lamellar Keratectomy | Stephen J. Koons | 09/265025<br>03/08/1999 | 6582445<br>06/24/2003 | |
| 018158-000800US | METHOD & APPARATUS FOR TEMPORAL & SPATIAL BEAM INTEGRATION | Herrmann J. Glocker | 08/368799<br>01/04/1995 | 5646791<br>07/08/1997 | |
| 018158-000810US | Method and Apparatus for Temporial and Spatial Beam Integration | Herrmann J. Glocker | 08/845510<br>04/24/1997 | 5912775<br>06/15/1999 | |
| 018158-000900US | Ophthalmological Surgery Technique with Active Patient Data Card | Alan R. McMillen<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 07/666840<br>03/08/1991 | 6296634<br>10/02/2001 | |
| 018158-000910US | Ophthalmological Surgery Technique with Active Patient Data Card | Alan R. McMillen<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 08/904199<br>07/31/1997 | 6364873<br>04/02/2002 | |
| 018158-000920US | Ophthalmological Surgery Technique with Active Patient Data Card | Alan R. McMillen<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 09/137975<br>08/21/1998 | 6106513<br>08/22/2000 | |
| 018158-000930US | Ophthalmological Surgery Technique With Active Patient Data Card | Alan R. McMillen<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 10/071960<br>02/06/2002 | 6846310<br>01/25/2005 | US-2002-0091376-A1<br>07/11/2002 |
| 018158-001000US | Method & Apparatus for Combined Cylindrical and Spherical Eye Corrections | John S. Shimmick | 07/746446<br>08/16/1991 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-001010US | METHOD & APPARATUS FOR COMBINED CYLINDRICAL & SPHERICAL EYE CORRECTIONS | John S. Shimmick | 08/138552<br>10/15/1993 | 5713892<br>02/03/1998 | |
| 018158-001030US | Method and Apparatus for Combined Cylindrical and Spherical Eye Corrections | John S. Shimmick | 08/698971<br>08/16/1996 | | |
| 018158-001031US | Method and Apparatus for Combined Cylindrical and Spherical Eye Corrections | John S. Shimmick | 08/841001<br>04/18/1997 | 6056740<br>05/02/2000 | |
| 018158-001100US | Surgical Laser System Microscope with Separated Ocular and Objective Lenses | Herrmann J. Glocker | 09/105073<br>06/26/1998 | 6251101<br>06/26/2001 | |
| 018158-001110US | Surgical Laser System Microscope with Separated Ocular and Objective Lenses | Herrmann J. Glocker | 09/852388<br>05/09/2001 | 6562026<br>05/13/2003 | US-2002-0049429-A1<br>04/25/2002 |
| 018158-001200US | EYE TRACKING DEVICE USING CORNEAL MARGIN DETECTION | Ioan T. Oltean | 60/062038<br>10/10/1997 | | |
| 018158-001210US | Eye Tracking Device for Laser Eye Surgery Using Corneal Margin Detector | Ioan T. Oltean<br>John S. Shimmick<br>Terrance N. Clapham | 09/167957<br>10/06/1998 | 6299307<br>10/09/2001 | |
| 018158-001300US | Method and System for Removing an Epithelial Layer from a Cornea | Terrance N. Clapham | 09/022774<br>02/12/1998 | 6068625<br>05/30/2000 | |
| 018158-001400US | Method for Scanning Non-Overlapping Patterns of Laser Energy With Diffractive Optics | Vladimir Lemberg | 09/116648<br>07/16/1998 | 6193710<br>02/27/2001 | |
| 018158-001410US | Method and System for Scanning Non-Overlapping Patterns of Laser Energy With Diffractive Optics | Vladimir Lemberg | 09/537201<br>03/28/2000 | 6491686<br>12/10/2002 | US-2002-0111606-A1<br>08/15/2002 |
| 018158-001500US | Integrated Epithelial Removal Tool | Terrance N. Clapham | 09/086793<br>05/28/1998 | 6132421<br>10/17/2000 | |
| 018158-001600US | SYSTEMS & METHODS FOR CALIBRATING LASER ABLATION SYSTEMS | Terrance N. Clapham<br>Angelina Thompson<br>George F. Caudle | 60/064724<br>11/06/1997 | | |
| 018158-001610US | Systems and Methods for Calibrating Laser Ablations | Angelina Thompson<br>Terrance N. Clapham<br>George F. Caudle | 09/187058<br>11/05/1998 | 6195164<br>02/27/2001 | |
| 018158-001700US | OPHTHALMIC METHOD & APPARATUS FOR LASER SURGERY OF THE CORNEA | F. L'Esperance Jr. | 08/468895<br>06/06/1995 | 5807379<br>09/15/1998 | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-001710US | OPHTHALMALIC METHOD & APPARATUS FOR LASER SURGERY OF THE CORNEA | F. L`Esperance Jr. | 09/008360<br>01/16/1998 | | |
| 018158-001800US | METHOD & SYSTEM FOR LASER TREATMENT OF REFRACTIVE ERRORS USING OFFSET IMAGING | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 08/058599<br>05/07/1993 | | |
| 018158-001810US | Method and System for Laser Treatment of Refractive Errors Using Offset Imaging | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 08/968380<br>11/12/1997 | 6203539<br>03/20/2001 | |
| 018158-001820US | METHOD & SYSTEM FOR LASER TREATMENT OF REFRACTIVE ERRORS USING OFFSET IMAGING | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 09/185914<br>11/04/1998 | | |
| 018158-001830US | Method and System for Laser Treatment of Refractive Errors Using Offset Imaging | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glockler | 09/730072<br>12/05/2000 | 6755818<br>06/29/2004 | US-2001-<br>0000530-A1<br>04/26/2001 |
| 018158-001840US | Method & System for Laser Treatment of Refractive Errors Using Offset Imaging | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 09/923863<br>08/06/2001 | | 0016586<br>02/07/2002 |
| 018158-001850US | Method and System for Laser Treatment of Refractive Errors Using Offset Imaging | William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glockler<br>John K. Shimmick | 10/831709<br>04/23/2004 | | US-2004-<br>0199224-A1<br>10/07/2004 |
| 018158-001900US | Rectilinear Photokeratoscope | John S. Shimmick<br>Charles R. Munnerlyn | 07/786650<br>11/01/1991 | 5339121<br>08/16/1994 | |
| 018158-001910US | RECTILINEAR PHOTOKERATOSCOPE | John S. Shimmick<br>Charles R. Munnerlyn | 08/184803<br>01/24/1994 | | |
| 018158-002000US | IN SITU ASTIGMATISM AXIS ALIGNMENT | John S. Shimmick<br>Charles R. Munnerlyn | 08/058733<br>05/07/1993 | | |
| 018158-002010US | In SITU Astigmatism Axis Alignment | John S. Shimmick<br>Charles R. Munnerlyn | 08/308480<br>09/19/1994 | 5549597<br>08/27/1996 | |
| 018158-002100US | METHOD & SYSTEM FOR TREATMENT OF REFRACTIVE ERROR USING AN OFFSET IMAGE OF A ROTATABLE MASK | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 08/058600<br>05/07/1993 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-002110US | Method & System for Laser Treatment of Refractive Error Using an Offset Image of a Rotatable Mask | John S. Shimmick<br>William B. Telfair<br>Charles R. Munnerlyn<br>Herrmann J. Glocker | 08/328937<br>10/25/1994 | 5556395<br>09/17/1996 | |
| 018158-003300US | METHOD & APPARATUS FOR PRECISION LASER SURGERY | H. Alfred Sklar<br>Alan M. Frank<br>Olga M. Ferrer<br>Charles F. McMillan<br>Stewart A. Brown<br>Fred Rienecker<br>Paul Harriss | 07/307315<br>02/06/1989 | 5098426<br>03/24/1992 | |
| 018158-003310US | METHOD & APPARATUS FOR PRECISION LASER SURGERY | H. Alfred Sklar<br>Alan M. Frank<br>Olga M. Ferrer<br>Charles F. McMillan<br>Stewart A. Brown<br>Fred Rienecker<br>Paul Harriss | 07/475657<br>02/06/1990 | | |
| 018158-003500US | SPECTRAL DIVISION OF REFLECTED LIGHT IN COMPLEX OPTICAL DIAGNOSTIC & THERAPEUTIC SYSTEMS | H. Alfred Sklar<br>Alan M. Frank | 07/523799<br>05/15/1990 | 5048946<br>09/17/1991 | |
| 018158-003510US | SPECTRAL DIVISION OF REFLECTED LIGHT IN COMPLEX OPTICAL DIAGNOSTIC & THERAPEUTIC SYSTEMS | H. Alfred Sklar<br>Alan M. Frank | 07/719924<br>06/24/1991 | 5157428<br>10/20/1992 | |
| 018158-004000US | Method of Establishing a Unique Machine Independent Reference Frame For the Eye | Carl F. Knopp<br>Paul R. Yoder Jr. | 08/091670<br>07/14/1993 | 5474548<br>12/12/1995 | |
| 018158-004100US | SYSTEM FOR SCANNING A SURGICAL LASER BEAM | William D. Fountain<br>Carl F. Knopp | 07/833604<br>08/22/1990 | 5391165<br>02/21/1995 | |
| 018158-004200US | SYSTEM & METHOD FOR DETECTING, CORRECTING, & MEASURING DEPTH MOVEMENT OF TARGET TISSUE IN A LASER SURGICAL SYSTEM | William D. Fountain | 07/655919<br>02/19/1991 | 5162641<br>11/10/1992 | |
| 018158-004210US | SYSTEM FOR DETECTING, CORRECTING & MEASURING DEPTH MOVEMENT OF A TARGET | William D. Fountain | 07/945207<br>09/15/1992 | 5286964<br>02/15/1994 | |
| 018158-004510US | APPARATUS & METHOD OF IDENTIFYING SIGNALS IN BIOLOGICAL TISSUES | Charles F. McMillan<br>H. Alfred Sklar | 07/656722<br>02/19/1991 | 5170193<br>12/08/1992 | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-004520US | ILLUMINATION OF THE CORNEA FOR PROFILOMETRY | Charles F. McMillan<br>William D. Fountain<br>Carl F. Knopp | 07/842879<br>02/27/1992 | 5283598<br>02/01/1994 | |
| 018158-004920US | AUTOMATED LASER WORKSTATION FOR HIGH PRECISION SURGICAL AND INDUSTRIAL INTERVENTIONS | Carl F. Knopp<br>William D. Fountain<br>Jerzy Orkiszewski<br>Michael Persiantsev<br>H. Alfred Sklar<br>Jan A. Wysopal | 07/843374<br>02/27/1992 | | |
| 018158-004921US | SYSTEM FOR DETECTING, MEASURING & COMPENSATING FOR LATERAL MOVEMENTS OF A TARGET | Carl F. Knopp<br>Jerzy Orkiszewski<br>Jan A. Wysopal<br>Hanna J. Hoffman | 08/019550<br>02/19/1993 | | |
| 018158-004922US | SYSTEM FOR DETECTING, MEASURING & COMPENSATING FOR LATERAL MOVEMENTS OF A TARGET | Carl F. Knopp<br>Jerzy Orkiszewski<br>Jan A. Wysopal<br>Hanna J. Hoffman | 08/515152<br>08/15/1995 | 5865832<br>02/02/1999 | |
| 018158-004930US | Automated Laser Workstation for High Precision Surgical and Industrial Interventions | Carl F. Knopp<br>William D. Fountain<br>Jerzy Orkiszewski<br>Michael Persiantsev<br>H. Alfred Sklar<br>Jan A. Wysopal | 08/523738<br>09/05/1995 | 6099522<br>08/08/2000 | |
| 018158-004940US | System for Detecting, Measuring and Compensating for Lateral Movements of a Target | Carl F. Knopp<br>Jerzy Orkiszewski<br>Jan A. Wysopal<br>Hanna J. Hoffman | 09/172976<br>10/14/1998 | 6702809<br>03/09/2004 | |
| 018158-004960US | Automated Laser Workstation for High Precision Surgical and Industrial Interventions | Carl F. Knoff<br>William D. Fountain<br>Jerzy Orkiszewski<br>Michael Persiantsev<br>H. Alfred Sklar<br>Jan A. Wysopal | 09/543840<br>04/05/2000 | 6726680<br>04/27/2004 | |
| 018158-004970US | Automated Laser Workstation for High Precison Surgical and Industrial Interventions | Carl F. Knopp<br>William D. Fountain<br>Jerzy Orkiszewski<br>Michael Persiantsev<br>H. Alfred Sklar<br>Jan A. Wysopal | 10/125261<br>04/17/2002 | 6913603<br>07/05/2005 | US-2002-0198516-A1<br>12/26/2002 |
| 018158-004980US | Automated Laser Workstation for High Precision Surgical and Industrial Interventions | Carl F. Knopp | 10/124891<br>04/17/2002 | | US-2002-0173778-A1<br>11/21/2002 |
| 018158-004990US | Automated Laser Workstation for High Precision Surgical and Industrial Interventions | Carl F. Knoff<br>William D. Fountain<br>Jerzy Orkiszewski<br>Michael Persiantsev<br>H. Alfred Sklar<br>Jan A. Wysopal | 10/632462<br>08/01/2003 | | US-2004-0059321-A1<br>03/25/2004 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-008000US | Linear Array Eye Tracker | Kingman Yee | 09/063879<br>04/21/1998 | 5966197<br>10/12/1999 | |
| 018158-008010US | Linear Array Eye Tracker | Kingman Yee | 09/249912<br>02/12/1999 | | |
| 018158-008020US | Linear Array Eye Tracker | Kingman Yee | 09/365428<br>08/02/1999 | 6283954<br>09/04/2001 | |
| 018158-008100US | Electromagnetic Drive System & Method for Gas Laser Fans | David I. Cullumber | 09/145819<br>09/02/1998 | | |
| 018158-008200US | Systems & Methods for Corneal Surface Ablation to Correct Hyperopia | Richard A. Hofer<br>Stephen J. Koons | 08/906020<br>08/05/1997 | | |
| 018158-008210US | Systems and Methods for Corneal Surface Ablation to Correct Hyperopia | Richard A. Hofer<br>Stephen J. Koons<br>John S. Shimmick | 09/379372<br>08/23/1999 | 6319247<br>11/20/2001 | |
| 018158-008220US | Systems and Methods for Corneal Surface Ablation to Correct Hyperopia | Richard A. Hofer<br>Stephen J. Koons<br>John S. Shimmick | 09/950563<br>09/10/2001 | 7582081<br>09/01/2009 | US-2002-0045889-A1<br>04/18/2002 |
| 018158-008600US | Laser Delivery System and Method With Diffractive Optic Beam Integration | George F. Caudle<br>Vladimir Lemberg | 09/015841<br>01/29/1998 | | |
| 018158-010400US | APPARATUS FOR OPHTHALMOLOGICAL SURGERY | F. L`Esperance Jr. | 06/916646<br>10/08/1986 | 4718418<br>01/12/1988 | |
| 018158-010700US | Patient Fixation System and Method for Laser Eye Surgery | John S. Shimmick<br>Vladimir Lemberg<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 09/105690<br>06/26/1998 | 6004313<br>12/21/1999 | |
| 018158-010710US | Patient Fixation System and Method for Laser Eye Surgery | John S. Shimmick<br>Vladimir Lemberg<br>Terrance N. Clapham<br>Charles R. Munnerlyn | 09/411785<br>10/01/1999 | 6406473<br>06/18/2002 | |
| 018158-010900US | MULTIPLE BEAM LASER SCULPTING SYSTEM & METHOD | Audrey Munnerlyn<br>George F. Caudle | 60/082156<br>04/17/1998 | | |
| 018158-010910US | Multiple Beam Laser Sculpting System and Method | Audrey Munnerlyn<br>George F. Caudle | 09/274499<br>03/23/1999 | 6331177<br>12/18/2001 | |
| 018158-010920US | Multiple Beam Laser Sculpting System & Method | Audrey Munnerlyn<br>George F. Caudle | 09/935509<br>08/22/2001 | 6638271<br>10/28/2003 | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-010930US | Multiple Beam Laser Sculpting System and Method | Audrey Munnerlyn<br>George F. Caudle | 10/327486<br>12/19/2002 | 6984227<br>01/10/2006 | US-2003-0125718-A1<br>07/03/2003 |
| 018158-011000US | REEXAMINATION OF U.S. PATENT NO. 4,903,695 | John Warner<br>F. L'Esperance Jr. | 90/004888<br>01/09/1998 | | |
| 018158-011100US | Method and Systems For Laser Treatment Of Presbyopia Using Offset Imaging | Marc Odrich<br>Kenneth Greenberg<br>Jerome A. Legerton<br>Charles R. Munnerlyn<br>John S. Shimmick | 60/076786<br>03/04/1998 | | |
| 018158-011110US | Method and Systems for Laser Treatment of Presbyopia Using Offset Imaging | Marc Odrich<br>Kenneth Greenberg<br>Jerome A. Legerton<br>Charles R. Munnerlyn<br>John S. Shimmick | 09/261768<br>03/03/1999 | 6280435<br>08/28/2001 | |
| 018158-011120US | Method and Systems for Laser Treatment of Presbyopia Using Offset Imaging | Marc Odrich<br>Kenneth Greenberg<br>Jerome A. Legerton<br>Charles R. Munnerlyn<br>John S. Shimmick | 09/901964<br>07/09/2001 | 6663619<br>12/16/2003 | |
| 018158-011140US | Method and Systems for Laser Treatment of Presbyopia Using Offset Imaging | Marc Odrich<br>Kenneth Greenberg<br>Jerome A. Legerton<br>Charles R. Munnerlyn<br>John S. Shimmick | 10/600027<br>06/19/2003 | | US-2004-0054356-A1<br>03/18/2004 |
| 018158-011141US | Method and Systems for Laser Treatment of Presbyopia Using Offset Imaging | Marc Odrich<br>Kenneth Greenberg<br>Jerome A. Legerton<br>Charles R. Munnerlyn<br>John K. Shimmick | 12/436736<br>05/06/2009 | | US 2009-0216217 A1<br>08/27/2009 |
| 018158-011300US | Digital Pulse Width Modulator | Ioan T. Oltean | 09/251317<br>02/17/1999 | 6044113<br>03/28/2000 | |
| 018158-011400US | Method and System for Ablating Surfaces with Partially Overlapping Craters Having Consistena Curvature | John S. Shimmick<br>George F. Caudle<br>Kingman Yee<br>Stephen J. Koons | 09/303810<br>04/30/1999 | 6497701<br>12/24/2002 | US-2002-015878-A1<br>10/17/2002 |
| 018158-011500US | Motion Detector for Eye Ablative Laser Delivery Systems | Richard A. Hofer | 09/370457<br>08/09/1999 | 6773430<br>08/10/2004 | US-2002-0077622-A1<br>06/20/2002 |
| 018158-011510US | Motion Detector for Eye Ablative Laser Delivery Systems | Richard A. Hofer | 10/760112<br>01/16/2004 | | |
| 018158-011600US | Integrated Scanning and Ocular Tomography System and Method | John S. Shimmick | 60/392330<br>06/27/2002 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-011610US | Integrated Scanning and Ocular Tomography System and Method | John S. Shimmick | 10/601119<br>06/19/2003 | 7133137<br>11/07/2006 | US-2004-0021874-A1<br>02/05/2004 |
| 018158-011620US | Integrated Scanning and Ocular Tomography System and Method | John S. Shimmick | 11/421450<br>05/31/2006 | 7503916<br>03/17/2009 | US-2006-0206102-A1<br>09/14/2006 |
| 018158-011700US | Offset Ablation Profiles for Treatment of Irregular Astigmatism | Terrance N. Clapham | 09/287322<br>04/07/1999 | 6245059<br>06/12/2001 | |
| 018158-011710US | Offset Ablation Profiles for Treatment of Irregular Astigmatism | Terrance N. Clapham | 09/823416<br>03/29/2001 | 6572607<br>06/03/2003 | US-2001-0020163-A1<br>09/06/2001 |
| 018158-011720US | Offset Ablation Profiles for Treatment of Irregular Astigmatism | Terrance N. Clapham | 10/402478<br>03/27/2003 | 7004935<br>02/28/2006 | US-2003-0220631-A1<br>11/27/2003 |
| 018158-012200US | Uniform Large Area Ablation System and Method | John S. Shimmick | 09/440826<br>11/15/1999 | 6530916<br>03/11/2003 | |
| 018158-012400US | Method and Apparatus for Determining Characteristics of a Laser Beam Spot | Kingman Yee<br>Terrance N. Clapham | 09/395809<br>09/14/1999 | 6559934<br>05/06/2003 | |
| 018158-012410US | Method and Systems for Laser Calibration and Eye Tracker Camera Alignment | Seema Somani<br>Kingman Yee<br>John S. Shimmick | 10/131622<br>04/23/2002 | 6666855<br>12/23/2003 | US-2002-0198515-A1<br>12/26/2002 |
| 018158-012420US | Method and Apparatus for Determining Characteristics of a Laser Beam Spot | Kingman Yee<br>Terrance N. Clapham | 10/383445<br>03/06/2003 | 7001375<br>02/21/2006 | US-2003-0149426-A1<br>08/07/2003 |
| 018158-012430US | Methods and Systems for Laser Calibration and Eye Tracker Camera Alignment | Seema Somani<br>Kingman Yee<br>John S. Shimmick | 10/685253<br>10/13/2003 | 7001376<br>02/21/2006 | US-2004-0078031-A1<br>04/22/2004 |
| 018158-012440US | Methods and Systems for Laser Calibration and Eye Tracker Camera Alignment | Seema Somani<br>Kingman Yee<br>John S. Shimmick | 11/264785<br>10/31/2005 | 7238177<br>07/03/2007 | US-2006-0103839-A1<br>05/18/2006 |
| 018158-012450US | Methods and Systems for Laser Calibration and Eye Tracker Camera Alignment | Seema Somani<br>Kingman Yee<br>John S. Shimmick | 11/755194<br>05/30/2007 | 7456949<br>11/25/2008 | US-2007-0225692-A1<br>09/27/2007 |
| 018158-012460US | Methods and Systems for Laser Calibration and Eye Tracker Camera Alignment | Seema Somani<br>Kingman Yee<br>John S. Shimmick | 12/200386<br>08/28/2008 | 7652761<br>01/26/2010 | US-2009-0076487-A1<br>03/19/2009 |
| 018158-012600US | Temperature Actuated Positioning Device for Nonlinear Optical Elements | George F. Caudle | 60/252634<br>11/22/2000 | | |
| 018158-012610US | Temperature Actuated Positioning Device for Non-Linear Optical Elements | George F. Caudle | 10/053513<br>04/11/2002 | 6768576<br>07/27/2004 | US-2002-0111611-A1<br>08/15/2002 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-012620US | Temperature Actuated Positioning Device for Nonlinear Optical Elements | George F. Caudle | 10/108187<br>03/26/2002 | | |
| 018158-012700US | Interface for Laser Eye Surgery | Terrance N. Clapham<br>Michael Cowperthwaite<br>Richard A. Hofer<br>Erik Scramaglia | 60/128122<br>04/07/1999 | | |
| 018158-012710US | Interface for Laser Eye Surgery | Michael Cowperthwaite | 09/534849<br>03/28/2000 | 6558373<br>05/06/2003 | |
| 018158-012720US | Interface for Laser Eye Surgery | Terance N. Clapham<br>Michael Cowperthwaite<br>Richard A. Hofer<br>Erik Scramaglia | 10/227176<br>08/20/2002 | 7001373<br>02/21/2006 | US-2003-0004502-A1<br>01/02/2003 |
| 018158-012730US | Interface for Laser Eye Surgery | Terrance N. Clapham | 10/226867<br>08/20/2002 | | US-2003-0004500-A1<br>01/02/2003 |
| 018158-012740US | Interface for Laser Eye Surgery | Terrance N. Clapham | 10/226993<br>08/20/2002 | | US-2003-0009159-A1<br>01/09/2003 |
| 018158-013000US | Hydration & Topography Tissue Measurements for Laser Sclupting | John S. Shimmick<br>Charles R. Munnerlyn<br>George F. Caudle<br>Terrance N. Clapham | 60/146231<br>07/28/1999 | | |
| 018158-013010US | Hydration and Topography Tissue Measurements for Laser Sculpting | John S. Shimmick<br>Charles R. Munnerlyn<br>George F. Caudle<br>Terrance N. Clapham | 09/626732<br>07/27/2000 | 6592574<br>07/15/2003 | |
| 018158-013200US | Generating Scanning Spot Locations for Laser Eye Surgery | Kingman Yee | 60/189633<br>03/14/2000 | | |
| 018158-013210US | Generating Scanning Spot Locations for Laser Eye Surgery | Kingman Yee<br>Erik Gross | 09/805737<br>03/13/2001 | 6673062<br>01/06/2004 | US-2002-0035359-A1<br>03/21/2002 |
| 018158-013211US | Generating Scanning Spot Locations for Laser Eye Surgery | Kingman Yee<br>Erik Gross | 10/655151<br>09/03/2003 | 7008415<br>03/07/2006 | US-2004-0059398-A1<br>03/25/2004 |
| 018158-013400US | Two-Pivot Scanning for Laser Eye Surgery | Herrmann J. Glocker<br>Henry Price<br>Jeff Sobiech | 09/405457<br>09/24/1999 | 6488676<br>12/03/2002 | |
| 018158-013600US | Optical Feedback System for Vision Correction | Vladimir Lemberg | 09/470907<br>12/23/1999 | 6419671<br>07/16/2002 | |
| 018158-013610US | Optical Feedback System for Vision Correction | Vladimir Lemberg | 10/160211<br>05/28/2002 | 6793654<br>09/21/2004 | US-2002-0169441-A1<br>11/14/2002 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-016200US | Two Camera Off-Axis Eye Tracker for Laser Eye Surgery | Kingman Yee<br>Charles R. Munnerlyn | 60/158576<br>10/07/1999 | | |
| 018158-016210US | Two Camera Off-Axis Eye Tracker for Laser Eye Surgery | Kingman Yee<br>Charles R. Munnerlyn | 09/545240<br>04/07/2000 | 6322216<br>11/27/2001 | |
| 018158-017200US | Flexible Scanning Beam Imaging System | George F. Caudle<br>Vladimir Lemberg | 60/356840<br>02/12/2002 | | |
| 018158-017210US | Flexible Scanning Beam Imaging System | George F. Caudle<br>Vladimir Lemberg | 10/367015<br>02/12/2003 | 7108691<br>09/19/2006 | US-2004-0073200-A1<br>04/15/2004 |
| 018158-017300US | Smoothing Laser Beam Integration Using Optical Element Motion | George F. Caudle<br>Herrmann J. Glockler<br>Wayne Brewer | 60/356720<br>02/12/2002 | | |
| 018158-017310US | Smoothing Laser Beam Integration Using Optical Element Motion | George F. Caudle<br>Herrmann J. Glockler<br>Wayne Brewer | 10/366131<br>02/12/2003 | 6816316<br>11/09/2004 | US-2004-0042080-A1<br>03/04/2004 |
| 018158-018600US | Direct Wavefront-Based Corneal Ablation Treatment Program | Lawrence W. Stark<br>John S. Shimmick | 60/254313<br>12/08/2000 | | |
| 018158-018610US | Direct Wavefront-Based Corneal Ablation Treatment Program | Lawrence W. Stark<br>John S. Shimmick | 10/006992<br>12/06/2001 | | US-2002-0135736-A1<br>09/26/2002 |
| 018158-018900US | Enhanced Wavefront Ablation System | John S. Shimmick<br>Jason Watson | 60/234452<br>09/21/2000 | | |
| 018158-018910US | Enhanced Wavefront Ablation System | Jason Watson<br>John S. Shimmick | 09/960163<br>09/20/2001 | 6685319<br>02/03/2004 | US-2002-0097375-A1<br>07/25/2002 |
| 018158-019400US | Application of Blend Zones, Depth Reduction and Transistion to Zones to Ablation Shapes | Erik Gross<br>Richard A. Hofer<br>Jonathan Wong | 10/100231<br>03/14/2002 | | US-2003-0176855-A1<br>09/18/2003 |
| 018158-019900US | Beam Position Monitoring for Laser Eye Surgery | Russell Schroeder | 10/128130<br>04/22/2002 | 6864478<br>03/08/2005 | US-2003-0197908-A1<br>10/23/2003 |
| 018158-020000US | Ablation Shape for the Correction of Presbyopia | Erik Gross | 10/300721<br>11/19/2002 | 6932808<br>08/23/2005 | US-2004-0097909-A1<br>05/20/2004 |
| 018158-020300US | Closed Loop System & Method for Ablating Lenses with Aberrations | Dimitri Chernyak<br>Charles E. Campbell<br>Jeffrey Jonathan Persoff | 60/356672<br>02/11/2002 | | |
| 018158-020310US | Closed Loop System and Method for Testing a Performance of a Laser System | Dimitri Chernyak<br>Charles E. Campbell<br>Jeffrey Jonathan Persoff | 10/364886<br>02/11/2003 | | US-2003-0225399-A1<br>12/04/2003 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-020311US | Closed Loop System and Method for Ablating Lenses with Aberrations | Dimitri Chernyak<br>Charles E. Campbell<br>Jeffrey Jonathan Persoff | 12/353842<br>01/14/2009 | | US-2009-0125005-A1<br>05/14/2009 |
| 018158-020400US | Corneal Topography-Based Target Warping | Dimitri Chernyak | 60/389090<br>06/13/2002 | | |
| 018158-020410US | Corneal Topography-Based Target Warping | Dimitri Chernyak | 10/460060<br>06/11/2003 | 7083609<br>08/01/2006 | US-2004-0019346-A1<br>01/29/2004 |
| 018158-020420US | Corneal Topography-Based Target Warping System | Dimitri Chernyak | 11/096536<br>03/31/2005 | 7419485<br>09/02/2008 | US-2005-0171515-A1<br>08/04/2005 |
| 018158-020430US | Corneal Topography-Based Target Warping System | Dimitri Chernyak | 12/172114<br>07/11/2008 | | US-2008-0275434-A1<br>11/06/2008 |
| 018158-020500US | Variable Repetition Rate Firing Scheme for Refractive Laser Systems | Jonathan Wong | 60/384621<br>05/30/2002 | | |
| 018158-020510US | Variable Repetition Rate Firing Scheme for Refractive Laser Systems | Jonathan Wong | 10/447665<br>05/28/2003 | 7077838<br>07/18/2006 | US-2005-0102008-A1<br>05/12/2005 |
| 018158-021200US | Apparatus & Method for Determining Relative Positional & Rotational Offsets Between a First & Second Device | Dimitri Chernyak<br>Jeffrey Jonathan Persoff | 60/356658<br>02/11/2002 | | |
| 018158-021210US | Apparatus and Method for Determining Relative Positional and Rotational Offsets Between a First and Second Imaging Device | Dimitri Chernyak<br>Jeffrey Jonathan Persoff | 10/365121<br>02/11/2003 | 7040759<br>05/09/2006 | US-2003-0151720-A1<br>08/14/2003 |
| 018158-021300US | Method & Device for DCalibrating an Optical Wavefront System | Jeffrey Jonathan Persoff | 60/356657<br>02/11/2002 | | |
| 018158-021310US | Method and Device for Calibrating an Optical Wavefront System | Jeffrey Jonathan Persoff | 10/364973<br>02/11/2003 | 7213919<br>05/08/2007 | US-2003-0169402-A1<br>09/11/2003 |
| 018158-021500US | Tracking a Torsional Position of an Eye | Dimitri Chernyak | 60/384653<br>05/30/2002 | | |
| 018158-021510US | Methods and Systems for Tracking a Torsional Orientation and Position of an Eye | Dimitri Chernyak | 10/300714<br>11/19/2002 | 7044602<br>05/16/2006 | US-2003-0223037-A1<br>12/04/2003 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-021520US | Methods and Systems for Tracking a Torsional Orientation and Position of an Eye | Dimitri Chernyak | 11/277743<br>03/28/2006 | 7261415<br>08/28/2007 | US-2006-0161141-A1<br>07/20/2006 |
| 018158-021530US | Methods and Systems for Tracking a Torsional Orientation and Position of an Eye | Dimitri Chernyak | 11/775840<br>07/10/2007 | 7431457<br>10/07/2008 | US-2008-0009840-A1<br>01/10/2008 |
| 018158-021540US | Method and Systems for Tracking a Torsional Orientation and Position of an Eye | Dimitri Chernyak | 12/210933<br>09/15/2008 | | US-2009-0012505-A1<br>01/08/2009 |
| 018158-021600US | Improved Hartmann-Shack Wavefront Measurement | Russell Schroder | 10/462102<br>06/12/2003 | | US-2004-0263779-A1<br>12/30/2004 |
| 018158-021700US | Eye Refractor With Active Mirror Wavefront Sensor | Charles E. Campbell | 10/364766<br>02/10/2003 | 6910770<br>06/28/2005 | |
| 018158-021710US | Eye Refractor With Active Mirror Wavefront Sensor | Charles E. Campbell | 11/059874<br>02/16/2005 | 7128416<br>10/31/2006 | US-2005-0286020-A1<br>12/29/2005 |
| 018158-021800US | Wavefront Reconstruction Using Fourier Transformation and Direct Integration | Dimitri Chernyak<br>Charles E. Campbell<br>Erik Gross<br>Seema Somani | 10/601048<br>06/20/2003 | 7175278<br>02/13/2007 | US-2004-0257530-A1<br>12/23/2004 |
| 018158-021810US | Iterative Fourier Reconstruction for Laser Surgery and Other Optical Applications | Dimitri Chernyak<br>Charles E. Campbell<br>Erik Gross<br>Seema Somani<br>Guangming Dai | 10/872107<br>06/17/2004 | 7168807<br>01/30/2007 | US-2005-0012898-A1<br>01/20/2005 |
| 018158-021820US | Iterative Fourier Reconstruction for Laser Surgery and Other Optical Applications | Dimitri Chernyak<br>Charles E. Campbell<br>Erik Gross<br>Seema Somani<br>Guangming Dai | 11/610937<br>12/14/2006 | 7365893<br>04/29/2008 | US-2007-0091263-A1<br>04/26/2007 |
| 018158-021830US | Iterative Fourier Reconstruction for Laser Surgery and Other Optical Applications | Dimitri Chernyak<br>Charles E. Campbell<br>Erik Gross<br>Seema Somani<br>Guangming Dai | 12/050651<br>03/18/2008 | 7731363<br>06/08/2010 | US-2008-0212031-A1<br>09/04/2008 |
| 018158-021840US | Iterative Fourier Reconstruction for Laser Surgery and Other Optical Applications | Dimitri Chernyak<br>Charles E. Campbell<br>Erik Gross<br>Seema Somani<br>Guangming Dai | 12/731638<br>03/25/2010 | | US-2010-0179793-A1<br>07/15/2010 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-021900US | Passive Gas Flow Management and Filtration Device for Use In an Excimer or Transverse Discharge Laser | Brian Bliven<br>David Turnquist | 10/632686<br>07/31/2003 | 6973112<br>12/06/2005 | US-2005-0025208-A1<br>02/03/2005 |
| 018158-022000US | Temperature Rise Model for Laser Vision Correction | Erik Gross<br>Kingman Yee<br>Jonathan Wong | 60/431355<br>12/06/2002 | | |
| 018158-022010US | Thermal Modeling for Reduction of Refractive Laser Surgery Times | Erik Gross<br>Kingman Yee<br>Jonathan Wong | 10/447667<br>05/28/2003 | 6964659<br>11/15/2005 | US-2004-0111083-A1<br>06/10/2004 |
| 018158-022100US | Methods for Determining Refractive Correction from Wavefront Aberration | Junzhong Liang | 60/431622<br>12/06/2002 | | |
| 018158-022110US | Methods for Determining Refractive Corrections from Wavefront Measurements | Junzhong Liang | 10/726733<br>12/02/2003 | 7460288<br>12/02/2008 | US-2004-0145702-A1<br>07/29/2004 |
| 018158-022200US | Shape Optimization for Presbyopia Correction | Guangming Dai | 60/431634<br>12/06/2002 | | |
| 018158-022210US | Shape Optimization for Presbyopia Correction | Guangming Dai | 60/468303<br>05/05/2003 | | |
| 018158-022220US | Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 10/738358<br>12/05/2003 | 7293873<br>11/13/2007 | US-2004-0169820-A1<br>09/02/2004 |
| 018158-022221US | Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 11/938648<br>11/12/2007 | 7520612<br>04/21/2009 | US-2008-0106698-A1<br>05/08/2008 |
| 018158-022222US | Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 12/207432<br>09/09/2008 | 7648238<br>01/19/2010 | US-2009-0066913-A1<br>03/12/2009 |
| 018158-022230US | Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 60/579124<br>06/10/2004 | | |
| 018158-022240US | Compound Modulation Transfer Function for Laser Surgery and Other Optical Applications | Guangming Dai<br>Kingman Yee | 10/911400<br>08/03/2004 | 7320517<br>01/22/2008 | US-2005-0254006-A1<br>11/17/2005 |
| 018158-022241US | Compound Modulation Transfer Function for Laser Surgery and Other Optical Applications | George M. Dai<br>Kingman Yee | 11/948475<br>11/30/2007 | 7475986<br>01/13/2009 | US-2008-0129962-A1<br>06/05/2008 |
| 018158-022242US | Compound Modulation Transfer Function For Laser Surgery And Other Optical Applications | George M. Dai<br>Kingman Yee | 12/329743<br>12/08/2008 | | US-2009-0086163-A1<br>04/02/2009 |
| 018158-022250US | Residual Accommodation Threshold for Correction of Presbyopia and Other Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 11/134630<br>05/19/2005 | 7434936<br>10/14/2008 | US-2005-0270491-A1<br>12/08/2005 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-022251US | Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 12/182062<br>07/29/2008 | | US-2009-0079940-A1<br>03/26/2009 |
| 018158-022260US | Residual Accomodation Threshold for Correction of Presbyopia and Other Presbyopia Correction Using Patient Data | Guangming Dai<br>Kingman Yee | 12/145862<br>06/25/2008 | 7762668<br>07/27/2010 | US-2008-030987-A1<br>12/18/2008 |
| 018158-022300US | Method to Optimize Laser Vision Correction of Presbyopia | Kingman Yee | 60/468387<br>05/05/2003 | | |
| 018158-022310US | Presbyopia Correction Using Effective Power | Guangming Dai<br>Kingman Yee | 60/519885<br>11/13/2003 | | |
| 018158-022500US | Wavefront Calibration Analyzer and Methods | Junzhong Liang<br>Dimitri Chernyak<br>Kingman Yee<br>Seema Somani<br>Jeffrey Jonathan Persoff<br>Walter Huff<br>Charles E. Campbell<br>Charles R. Munnerlyn | 60/461739<br>04/09/2003 | | |
| 018158-022510US | Treatment Plane Wavefront Calibration Analyzer and Methods | Junzhong Liang<br>Dimitri Chernyak<br>Kingman Yee<br>Seema Somani<br>Jeffrey Jonathan Persoff<br>Walter Huff<br>Charles E. Campbell<br>Charles R. Munnerlyn<br>Brian Bliven | 60/518867<br>11/10/2003 | | |
| 018158-022520US | Wavefront Calibration Analyzer and Methods | Junzhong Liang<br>Dimitri Chernyak<br>Kingman Yee<br>Seema Somani<br>Jeffrey Jonathan Persoff<br>Walter Huff<br>Charles E. Campbell<br>Charles R. Munnerlyn<br>Brian Bliven | 10/799439<br>03/12/2004 | 7355695<br>04/08/2008 | US-2004-0260275-A1<br>12/23/2004 |
| 018158-022600US | Methods and Devices for Registering Optical Measurement Datasets of an Optical System | Dimitri Chernyak | 10/463674<br>06/16/2003 | 7458683<br>12/02/2008 | US-2004-0263785-A1<br>12/30/2004 |
| 018158-022700US | Systems and Methods for Eye Aberration and Image Sensor Orientation | Jeffrey Jonathan Persoff | 60/491920<br>07/31/2003 | | |
| 018158-022710US | Systems and Methods for Eye Aberration and Image Sensor Orientation | Jeffrey Jonathan Persoff | 10/839792<br>05/04/2004 | 7338164<br>03/04/2008 | US-2005-0024584-A1<br>02/03/2005 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-022800US | Fiber Optic Slit Source for Ophthalmic Use | Seema Somani | 60/491801<br>08/01/2003 | | |
| 018158-022810US | Slit Lamp for Ophthalmic Use | Seema Somani | 10/872026<br>06/17/2004 | 7338169<br>03/04/2008 | US-2005-0024587-A1<br>02/03/2005 |
| 018158-022820US | Slit Lamp for Opthalmic Use | Seema Somani | 11/969619<br>01/04/2008 | 7431459<br>10/07/2008 | US-2008-0106699-A1<br>05/08/2008 |
| 018158-023000US | Systems and Methods for Correcting High Order Aberrations in Laser Refractive Surgery | Guangming Dai<br>Junzhong Liang | 60/463873<br>04/18/2003 | | |
| 018158-023010US | Systems and Methods for Correcting High Order Aberrations In Laser Refractive Surgery | Guangming Dai<br>Junzhong Liang | 10/825864<br>12/20/2004 | | US-2005-0096640-A1<br>05/05/2005 |
| 018158-023100US | Database System for Centralized Clinical and Research Applications With Data from Wavefront Aberrometers | Guangming Dai<br>Kingman Yee<br>Dimitri Chernyak | 60/644227<br>01/13/2005 | | |
| 018158-023110US | Database System for Centralized Clinical and Research Applications With Data from Wavefront Aberrometers | Guangming Dai<br>Kingman Yee<br>Dimitri Chernyak | 11/332824<br>01/12/2006 | | US-2006-0173644-A1<br>08/03/2006 |
| 018158-023200US | Lenslet Array for Beam Homogenization | Seema Somani<br>Charles R. Munnerlyn<br>Mark E. Arnoldussen<br>John Osborn | 10/913952<br>08/06/2004 | 7206132<br>04/17/2007 | US-2006-0028732-A1<br>02/09/2006 |
| 018158-023210US | Lenslet Array for Beam Homogenization | Seema Somani<br>Charles R. Munnerlyn<br>Mark E. Arnoldussen<br>John Osborn | 11/683963<br>03/08/2007 | 7394595<br>07/01/2008 | US-2007-0146890-A1<br>06/28/2007 |
| 018158-023220US | Lenslet Array for Beam Homogenization | Seema Somani<br>Charles R. Munnerlyn<br>Mark E. Arnoldussen<br>John Osborn | 11/683968<br>03/08/2007 | 7355794<br>04/08/2008 | US-2007-0146891-A1<br>06/28/2007 |
| 018158-023230US | Lenslet Array for Beam Homogenization | Seema Somani<br>Charles R. Munnerlyn<br>Mark E. Arnoldussen<br>John Osborn | 12/121475<br>05/15/2008 | 7738176<br>06/15/2010 | US-2009-0122411-A1<br>05/14/2009 |
| 018158-023300US | Systems and Methods for Prediction of Objective Visual Acuity Based on Wavefront Measurements | Guangming Dai | 60/480237<br>06/20/2003 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-023310US | Systems and Methods for Prediction of Objective Visual Acuity Based On Wavefront Measurements | Guangming Dai | 10/871344<br>06/18/2004 | 7338165<br>03/04/2008 | US-2005-0024585-A1<br>02/03/2005 |
| 018158-023311US | Systems and Methods for Prediction of Objective Visual Acuity Based on Wavefront Measurements | Guangming Dai | 12/013763<br>01/14/2008 | 7699470<br>04/20/2010 | US-2009-0021694-A1<br>01/22/2009 |
| 018158-023312US | Systems and Methods for Prediction of Objective Visual Acuity Based on WaveFront Measurements | Guangming Dai | 12/649575<br>12/30/2009 | | US-2010-0103376-A1<br>04/29/2010 |
| 018158-023400US | Sterile Hand Held Refractive Surgery Slit Lamp Illumination System | Cary Spediacci<br>John Weberg<br>Brian Bliven | 60/491909<br>08/01/2003 | | |
| 018158-023410US | Sterile Hand Held Refractive Surgery Slit Lamp Illumination System | Cary Spediacci<br>John Weberg<br>Brian Bliven | 10/876268<br>06/23/2004 | 7331672<br>02/19/2008 | US-2005-0099602-A1<br>05/12/2005 |
| 018158-023420US | Sterile Hand Held Refractive Surgery Slit Lamp Illumination System | Cary Spediacci<br>John Weberg<br>Brian Bliven | 11/959363<br>12/18/2007 | | US-2008-0094574-A1<br>04/24/2008 |
| 018158-023500US | Method and Apparatus for Measurement of the Refractive Properties of the Human Eye | Josef Bille<br>Frieder Loesel | 09/332297<br>06/11/1999 | 6050687<br>04/18/2000 | |
| 018158-023510US | Method and Apparatus for Measurement for Precompensating the Refractive Properties of the Human Eye With Adaptive Optical Feedback Control | Josef Bille<br>Frieder Loesel | 09/334774<br>06/16/1999 | 6155684<br>12/05/2000 | |
| 018158-023600US | Opthallmic Device for Mapping the Acuity of a Human Eye Using Wavefront Analysis | Wolfram Becker | 29/121211<br>04/03/2000 | D436665<br>01/23/2001 | |
| 018158-023800US | Microscope Magnification Sensor | Cary Spediacci<br>David Hindi<br>Mathew Clopp<br>Bill Fish | 10/703195<br>11/05/2003 | | US-2005-0094262-A1<br>05/05/2005 |
| 018158-024000US | Binocular Optical Treatment for Presbyopia | Dimitri Chernyak | 10/849573<br>05/18/2004 | | US-2005-0261752-A1<br>11/24/2005 |
| 018158-024020US | Presbyopia Correction Using Patient Data | Dimitri Chernyak<br>Guangming Dai<br>Kingman Yee | 12/474144<br>05/28/2009 | | US 2009-0234336 A1<br>09/17/2009 |
| 018158-024300US | Transformation Methods of Wavefront Maps from One Vertex Distance to Another | Guangming Dai | 60/550514<br>03/03/2004 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-024310US | Transformation Methods of Wavefront Maps from One Vertex Distance to Another | Guangming Dai | 11/032469<br>01/07/2005 | 7296893<br>11/20/2007 | US-2005-0195364-A1<br>09/08/2005 |
| 018158-024320US | Transformation Methods of Wavefront Maps from One Vertex Distance to Another | Guangming Dai | 11/871567<br>10/12/2007 | 7726813<br>06/01/2010 | US-2008-0252848-A1<br>10/16/2008 |
| 018158-024330US | Transformation Methods of Wavefront Maps from One Vertex Distance to Another | Guangming Dai | 12/786756<br>05/25/2010 | | |
| 018158-024500US | Calibrating Laser Beam Position and Shape Using an Image Capture Device | Dimitri Chernyak<br>Keith Holliday<br>Mathew Clopp | 10/808728<br>03/24/2004 | | US-2005-0215986-A1<br>09/29/2005 |
| 018158-024600US | Laser Pulse Position Monitor for Scanned Laser Eye Surgery Systems | David Hindi<br>Keith Holliday<br>Mark E. Arnoldussen | 60/586557<br>07/09/2004 | | |
| 018158-024610US | Laser Pulse Position Monitor for Scanned Laser Eye Surgery Systems | David Hindi<br>Keith Holliday<br>Mark E. Arnoldussen | 11/174279<br>06/30/2005 | 7479138<br>01/20/2009 | US-2006-0084955-A1<br>04/20/2006 |
| 018158-025100US | Methods and Systems for Differentiating Left and Right Eye Images | Jonathan Wong<br>Dimitri Chernyak | 10/784481<br>02/19/2004 | 7481536<br>01/27/2009 | US-2005-0185138-A1<br>08/25/2005 |
| 018158-025110US | Methods and Systems for Differentiating Left and Right Eye Images | Jonathan Wong<br>Dimitri Chernyak | 12/337956<br>12/18/2008 | | US-2009-0099558-A1<br>04/16/2009 |
| 018158-025200US | Methods and Devices for Testing Torsional Alignment Between a Diagnostic Device and a Laser Refractive System | Charles E. Campbell<br>Dimitri Chernyak | 60/518826<br>11/10/2003 | | |
| 018158-025210US | Methods and Devices for Testing Torsional Alignment Between a Diagnostic Device and a Laser Refractive System | Charles E. Campbell<br>Dimitri Chernyak | 10/985311<br>11/09/2004 | | US-2005-0131398-A1<br>06/16/2005 |
| 018158-025300US | Stabilizing Delivered Laser Energy | Keith Holliday | 60/553580<br>03/15/2004 | | |
| 018158-025310US | Stabilizing Delivered Laser Energy | Keith Holliday | 11/077173<br>03/09/2005 | | US-2005-0203494-A1<br>09/15/2005 |
| 018158-025400US | Correction of Presbyopia Using Adaptive Optics and Associated Methods | Guangming Dai | 10/872331<br>06/17/2004 | 7387387<br>06/17/2008 | US-2005-0280777-A1<br>12/22/2005 |
| 018158-025410US | Correction of Presbyopia Using Adaptive Optics and Associated Methods | Guangming Dai<br>Seema Somani | 11/156257<br>06/17/2005 | 7513620<br>04/07/2009 | US-2006-0017883-A1<br>01/26/2006 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-025411US | Correction of Presbyopia using Adaptive Optics, Wavefront Sensor Eye Alignment and Light Shield, and Associated Methods | Guangming Dai<br>Seema Somani | 12/335337<br>12/15/2008 | 7771048<br>08/10/2010 | US-2009-0122265-A1<br>05/14/2009 |
| 018158-025420US | Correction of Presbyopia Using Adaptive Optics and Associated Methods | Guangming Dai | 12/124116<br>05/20/2008 | 7708410<br>05/04/2010 | US-2008-0218693-A1<br>09/11/2008 |
| 018158-025430US | Correction of Presbyopia Using Adaptive Optics and Associated Methods | Guangming Dai | 12/177810<br>07/22/2008 | 7748847<br>07/06/2010 | US-2009-0015789-A1<br>01/15/2009 |
| 018158-025440US | Correction of Presbyopia Using Adaptive Optics, Wavefront Sensor Eye Alignment and Light Shield and Associated Methods | Guangming Dai<br>Seema Somani | 12/191510<br>08/14/2008 | 7690789<br>04/06/2010 | US-2008-0297723-A1<br>12/04/2008 |
| 018158-025500US | Volumetric Point Spread Function for Eye Diagnosis and Treatment | Erik Gross<br>Guangming Dai<br>Charles E. Campbell | 60/546416<br>02/20/2004 | | |
| 018158-025510US | Volumetric Point Spread Function for Eye Diagnosis and Treatment | Erik Gross<br>Guangming Dai<br>Charles E. Campbell | 60/548709<br>02/27/2004 | | |
| 018158-025520US | Volumetric Point Spread Function for Eye Diagnosis and Treatment | Erik Gross<br>Guangming Dai<br>Charles E. Campbell | 11/064876<br>02/22/2005 | 7377648<br>05/27/2008 | US-2005-0213040-A1<br>09/29/2005 |
| 018158-025800US | Sphero Cylindrical Eye Refraction System Using Fluid Focus Electrostatically Variable Lenses | Charles E. Campbell | 10/993409<br>11/18/2004 | 7413306<br>08/19/2008 | US-2006-0106426-A1<br>05/18/2006 |
| 018158-025810US | Sphero Cylindrical Eye Refraction System Using Fluid Focus Electrostatically Variable Lenses | Charles E. Campbell | 12/173756<br>07/15/2008 | | US-2008-0266521-A1<br>10/30/2008 |
| 018158-025900US | Scleral Lenses for Custom Optic Evaluation and Visual Performance Improvement | Kuang-Mon Ashley Tuan<br>Mark E. Arnoldussen<br>Guangming Dai<br>Kingman Yee | 11/134861<br>05/20/2005 | | US-2006-0264917-A1<br>11/23/2006 |
| 018158-026200US | Pupilometer for Pupil Center Drift and Pupil Size Measurements at Differing Viewing Distances | Dimitri Chernyak | 11/088010<br>03/22/2005 | 7431455<br>10/07/2008 | US-2006-0215113-A1<br>09/28/2006 |
| 018158-026210US | Pupilometer for Pupil Center Drift and Pupil Size Measurements at Differing Viewing Distances | Dimitri Chernyak | 12/197774<br>08/25/2008 | 7708405<br>05/04/2010 | US-2008-0309870-A1<br>12/18/2008 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-026211US | Pupilometer for Pupil Center Drift and Pupil Size Measurements at Differing Viewing Distances | Dimitri Chernyak | 12/731959<br>03/25/2010 | | |
| 018158-026300US | Presbyopia Correction Through Negative High-Order Spherical Aberration | Seema Somani<br>Kingman Yee | 11/173904<br>06/30/2005 | 7261412<br>08/28/2007 | US-2007-0002274-A1<br>01/04/2007 |
| 018158-026310US | Presbyopia Correction Through Negative High-Order Spherical Aberration | Seema Somani<br>Kingman Yee | 11/780147<br>07/19/2007 | 7478907<br>01/20/2009 | US-2008-0015461-A1<br>01/17/2008 |
| 018158-026320US | Presbyopia Correction Through Negative High-Order Spherical Aberration | Seema Somani<br>Kingman Yee | 12/207444<br>09/09/2008 | | US-2009-0000628-A1<br>01/01/2009 |
| 018158-026321US | Presbyopia Correction Through Negative Spherical Aberration | Seema Somani<br>Kingman Yee | 12/436743<br>05/06/2009 | | US 2009-0216218 A1<br>08/27/2009 |
| 018158-026400US | Training Enhanced Pseudo Accomodation Methods, Systems and Devices for Mitigation of Presbyopia | Kingman Yee | 11/134027<br>05/19/2005 | 7413566<br>08/19/2008 | US-2006-0264916-A1<br>11/23/2006 |
| 018158-026410US | Training Enhanced Pseudo Accommodation Methods, Systems and Devices for Mitigation of Presbyopia | Kingman Yee | 12/172127<br>07/11/2008 | | US-2009-0036981-A1<br>02/05/2009 |
| 018158-026900US | Scaling Zernike Coefficients to Smaller Pupil Sizes for Refractive Treatments | Guangming Dai | 60/776289<br>02/24/2006 | | |
| 018158-026910US | Scaling Zernike Coefficients to Smaller Pupil Sizes for Refractive Treatments | Guangming Dai | 11/676094<br>02/16/2007 | 7717562<br>05/18/2010 | US-2007-0201001-A1<br>08/30/2007 |
| 018158-026920US | Scaling Zernike Coefficients to Smaller Pupil Sizes for Refractive Treatments | Guangming Dai | 12/722881<br>03/12/2010 | | US-2010-0198567-A1<br>08/05/2010 |
| 018158-027200US | Sterile Hand Held Slit Lamp Cover and Method | Amy Keller<br>John Weberg | 11/123962<br>05/06/2005 | 7422328<br>09/09/2008 | US-2006-0250796-A1<br>11/06/2006 |
| 018158-027300US | Sterile Cover for Doctor`s Console for Laser Eye Surgery | John Weberg<br>Melinda J. Lennon<br>Amy Keller<br>Mark Brinkerhoff | 29/230907<br>05/26/2005 | D553749<br>10/23/2007 | |
| 018158-027400US | Chair Stabilizer for Refractive Surgery | Mark Brinkerhoff<br>Tom Kowalski<br>Shandon Alderson | 11/342278<br>01/27/2006 | 7661161<br>02/16/2010 | US-2007-0174971-A1<br>08/02/2007 |
| 018158-027410US | Chair Stabilizer for Refractive Surgery | Mark Brinkerhoff<br>Tom Kowalski<br>Shandon Alderson | 12/683580<br>01/07/2010 | | US-2010-0095455-A1<br>04/22/2010 |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-027500US | Compression Head Pillows and Neck Angle Adjustment Mechanism for Refractive Laser and the Like | Mark Brinkerhoff<br>Tom Kowalski<br>Shandon Alderson | 11/335177<br>01/18/2006 | 7451507<br>11/18/2008 | US-2007-0163049-A1<br>07/19/2007 |
| 018158-027600US | Laser Energy Calibration Based On Optical Measurement | Leander Zickler | 11/341917<br>01/26/2006 | | US-2007-0173797-A1<br>07/26/2007 |
| 018158-027700US | Output Energy Control For Lasers | Keith Holliday | 11/373069<br>03/10/2006 | | US-2007-0213697-A1<br>09/13/2007 |
| 018158-027800US | Spot Spacing Estimation and Refraction Calculation for Hartmann-Shack Waverfront Sensors | Erik Gross<br>Charles E. Campbell | 60/782679<br>03/14/2006 | | |
| 018158-027810US | Spatial Frequency Wavefront Sensor System and Method | Erik Gross<br>Charles E. Campbell | 11/686226<br>03/14/2007 | 7652235<br>01/26/2010 | US-2008-0073525-A1<br>03/27/2008 |
| 018158-027820US | Spatial Frequency Wavefront Sensor System and Method | Erik Gross<br>Charles E. Campbell | 12/641011<br>12/17/2009 | | US-2010-0090090-A1<br>04/15/2010 |
| 018158-028000US | Calculating Zernicke Coefficients from Fourier Coefficients | Guangming Dai | 11/218833<br>09/02/2005 | 7331674<br>02/19/2008 | US-2007-0058132-A1<br>03/15/2007 |
| 018158-028010US | Calculating Zernike Coefficients from Fourier Coefficients | Guangming Dai | 11/960094<br>12/19/2007 | 7748848<br>07/06/2010 | US-2008-0140329-A1<br>06/12/2008 |
| 018158-028200US | Non-Invasive Measurement of Tear Volume Systems and Methods | Kuang-Mon Ashley Tuan | 11/335861<br>01/18/2006 | 7520609<br>04/21/2009 | US-2007-0171365-A1<br>07/26/2007 |
| 018158-028210US | Non-Invasive Measurement of Tear Volume Systems and Methods | Kuang-Mon Ashley Tuan | 12/403229<br>03/12/2009 | | US-2009-0168019-A1<br>07/02/2009 |
| 018158-028300US | Systems and Methdos for Qualifying and Calibrating a Beam Delivery System | Mark E. Arnoldussen | 11/339984<br>01/25/2006 | | US-2007-0173792-A1<br>07/26/2007 |
| 018158-028500US | Shack-Hartmann Based Integrated Autorefraction and Wavefront Measurements of the Eye | Charles E. Campbell<br>Seema Somani | 11/376725<br>03/14/2006 | 7475989<br>01/13/2009 | US-2007-0216867-A1<br>09/20/2007 |
| 018158-028600US | Intrastromal Refractive Correction Systems and Methods | Kingman Yee | 60/783306<br>03/17/2006 | | |
| 018158-028601US | Intrastromal Refractive Correction Systems and Methods | Kingman Yee | 60/743535<br>03/17/2006 | | |

| TTC Ref No.<br>Attys Handling | Title | Inventor | App No.<br>Filing Date | Pat No.<br>Issue Date | Pub No.<br>Pub Date |
|---|---|---|---|---|---|
| 018158-028610US | Intrastromal Refractive Correction Systems and Methods | Kingman Yee | 11/677504<br>02/21/2007 | | US-2007-0219543-A1<br>09/20/2007 |
| 018158-028611US | Intrastromal Refractive Correction Systems and Methods | Kingman Yee | 12/471090<br>05/22/2009 | | US 2009-0234335 A1<br>09/17/2009 |
| 018158-028900US | Wavefront Propagation from One Plane to Another | Guangming Dai | 60/826636<br>09/22/2006 | | |
| 018158-028910US | Wavefront Propagation from One Plane to Another | Guangming Dai | 11/736353<br>04/17/2007 | 7547102<br>06/16/2009 | US-2007-0211214-A1<br>09/13/2007 |
| 018158-028911US | Wavefront Propagation from One Plane to Another | Guangming Dai<br>Charles E. Campbell<br>Li Chen<br>Huawei Zhao<br>Dimitri Chernyak | 12/421246<br>04/09/2009 | | US 2009-0231546 A1<br>09/17/2009 |
| 018158-029000US | Systems and Methods for Wavefront Reconstruction for Aperture with Arbitrary Shape | Guangming Dai | 60/785967<br>03/23/2006 | | |
| 018158-029010US | Systems and Methods for Wavefront Reconstruction for Aperture with Arbitrary Shape | Guangming Dai | 11/690409<br>03/23/2007 | 7780294<br>08/24/2010 | US-2007-0222948-A1<br>09/27/2007 |
| 018158-029020US | Systems and Methods for Wavefront Reconstruction for Aperture with Arbitrary Shape | Guangming Dai | 12/793348<br>06/03/2010 | | |
| 018158-029300US | System and Method for Illumination and Fixation with Ophthalmic Diagnostic Instruments | Kingman Yee<br>Seema Somani | 11/750291<br>05/17/2007 | | US-2008-0284979-A1<br>11/20/2008 |
| 018158-029400US | Auto-Alignment and Auto-Focus System and Method | Seema Somani | 11/747582<br>05/11/2007 | 7575322<br>08/18/2009 | US-2008-0278687-A1<br>11/13/2008 |
| 018158-030000US | Customized Laser Epithelial Ablation System and Methods | Keith Holliday<br>Mark E. Arnoldussen | 60/938684<br>05/17/2007 | | |
| 018158-030100US | Operator-Controlled Scanning Laser Procedure Designed for Large-Area Epithelium Removal | Mark E. Arnoldussen<br>Jonathan Wong<br>Benjamin A. Logan | 60/865342<br>11/10/2006 | | |
| 018158-030110US | Operator-Controlled Scanning Laser Procedure Designed for Large-Area Epithelium Removal | Mark E. Arnoldussen<br>Jonathan Wong<br>Benjamin A. Logan | 11/937760<br>11/09/2007 | | US-2008-0125763-A1<br>05/29/2008 |
| 018158-030120US | Operator-Controlled Scanning Laser Procedure Designed for Large-Area Epithelium Removal | Mark E. Arnoldussen<br>Jonathan Wong<br>Benjamin A. Logan<br>Leander Zickler | 12/121635<br>05/15/2008 | | US-2008-0287928-A1<br>11/20/2008 |

# APPENDIX B

# United States Patent [19]

**Sklar et al.**

[11] **Patent Number: 5,054,907**

[45] **Date of Patent: Oct. 8, 1991**


[54] **OPHTHALMIC DIAGNOSTIC APPARATUS AND METHOD**


[75] Inventors: **H. Alfred Sklar**, San Francisco; **Alan M. Frank**; **Charles McMillan**, both of Livermore, all of Calif.; **Olga M. Ferrer**, Miami, Fla.

[73] Assignee: **Phoenix Laser Systems, Inc.**, San Francisco, Calif.



[21] Appl. No.: **456,109**

[22] Filed: **Dec. 22, 1989**

[51] **Int. Cl.**[5] ..... **A61B 3/10**

[52] **U.S. Cl.** ..... **351/212**; 351/247

[58] **Field of Search** ..... 351/212, 247; 356/376



[56] **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| 4,312,574 | 1/1982 | Wilms | 351/212 |
| 4,878,750 | 11/1989 | Sekiguchi | 351/212 |

OTHER PUBLICATIONS

S. G. El Hage, "Suggested New Methods for Photokeratoscopy, A Comparison of their Validities, Part I", *American Journal of Optometry and Archives of American Academy of Optometry*, pp. 897-912, 11/71.



*Primary Examiner*—Rodney B. Bovernick

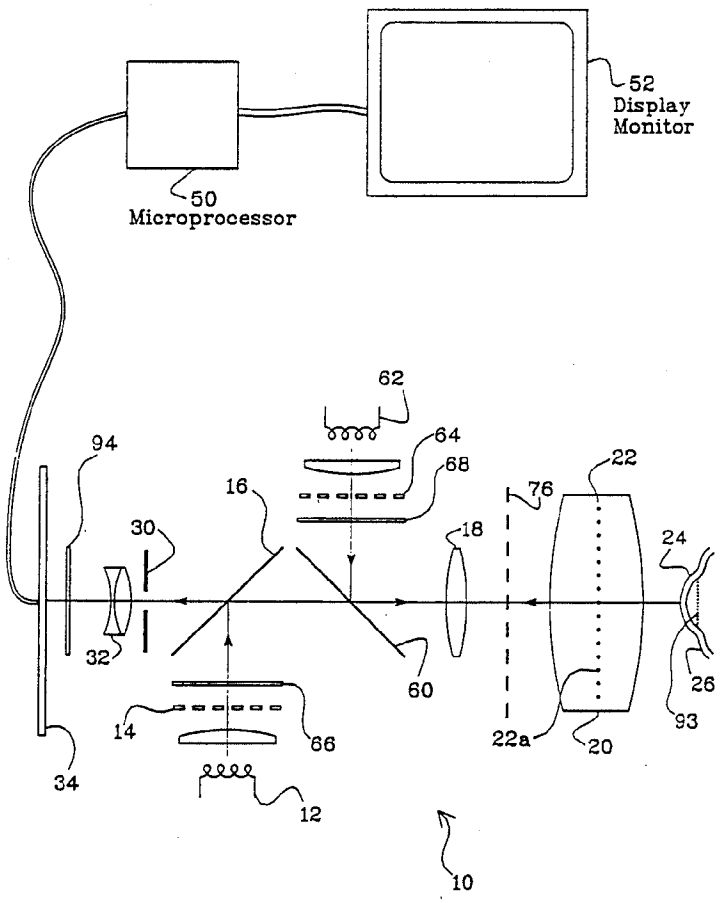
*Attorney, Agent, or Firm*—Thomas M. Freiburger



[57] **ABSTRACT**

An ophthalmic diagnostic instrument determines the shape of the cornea through projection of an image onto the cornea through the optics of the diagnostic instrument. The instrument and the method of the invention involve folding a projected pattern of discrete separated point light sources so that the pattern is projected toward the eye coaxially with return collected light reflected off the cornea. The instrument avoids any need for a pattern light source directly adjacent to the eye, and provides the surgeon or other eye care specialist with a real time image accurately displaying the shape of the cornea. The surgeon is thus able to monitor the corneal shape prior to surgery, to monitor its changes during the course of the surgery, and to further monitor the cornea in post operative stages. In a specific embodiment of the invention, a real image of the pattern of point light sources is formed inside or very closely in front of the objective lens of the system so that the objective lens becomes a field lens and the angle of view of the system is enlarged.


**30 Claims, 8 Drawing Sheets**



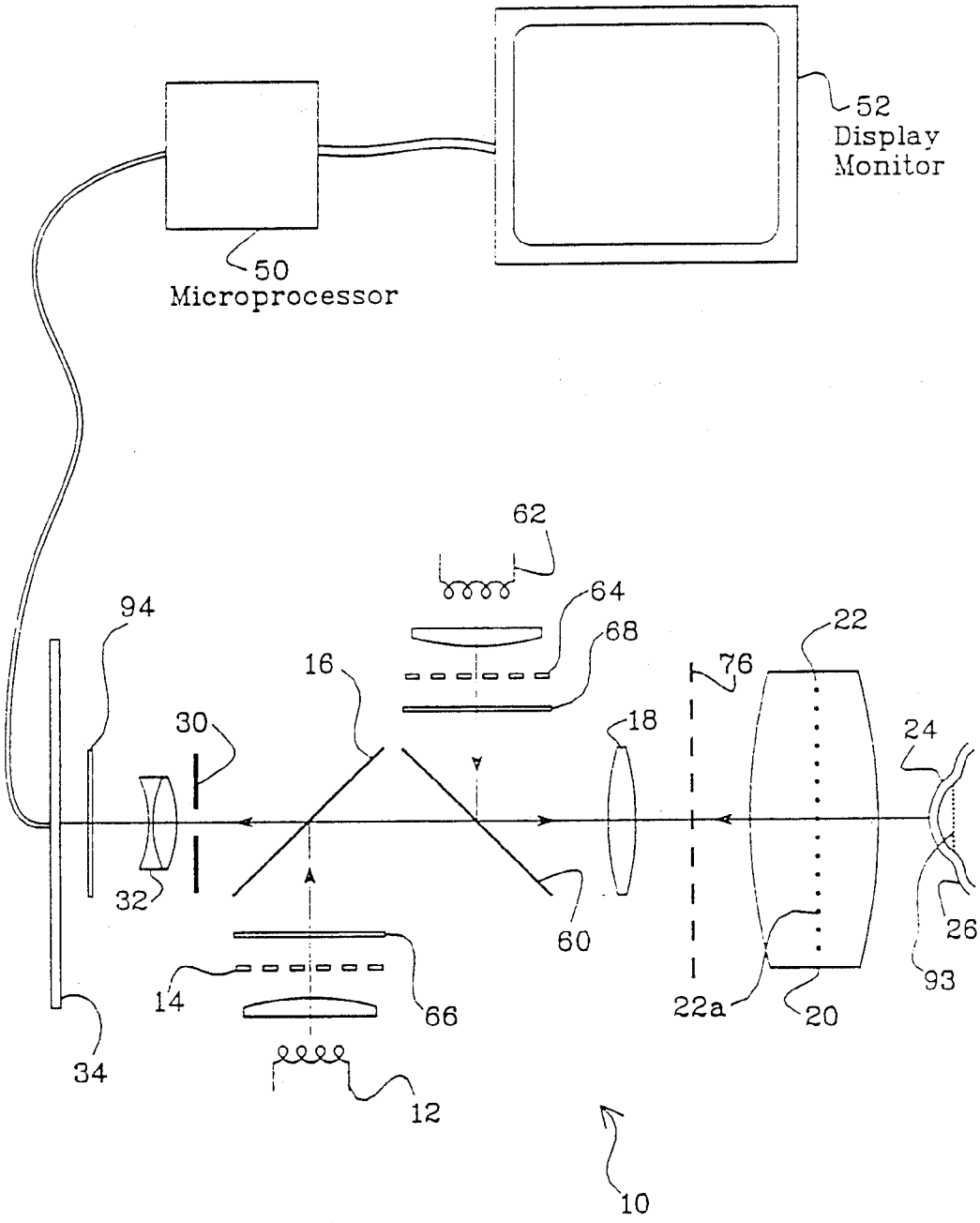

52
Display
Monitor
50
Microprocessor
62
64
68
94
16
22
76
30
18
24
32
60
26
93
14
66
22a
20
34
12
10

Figure 1

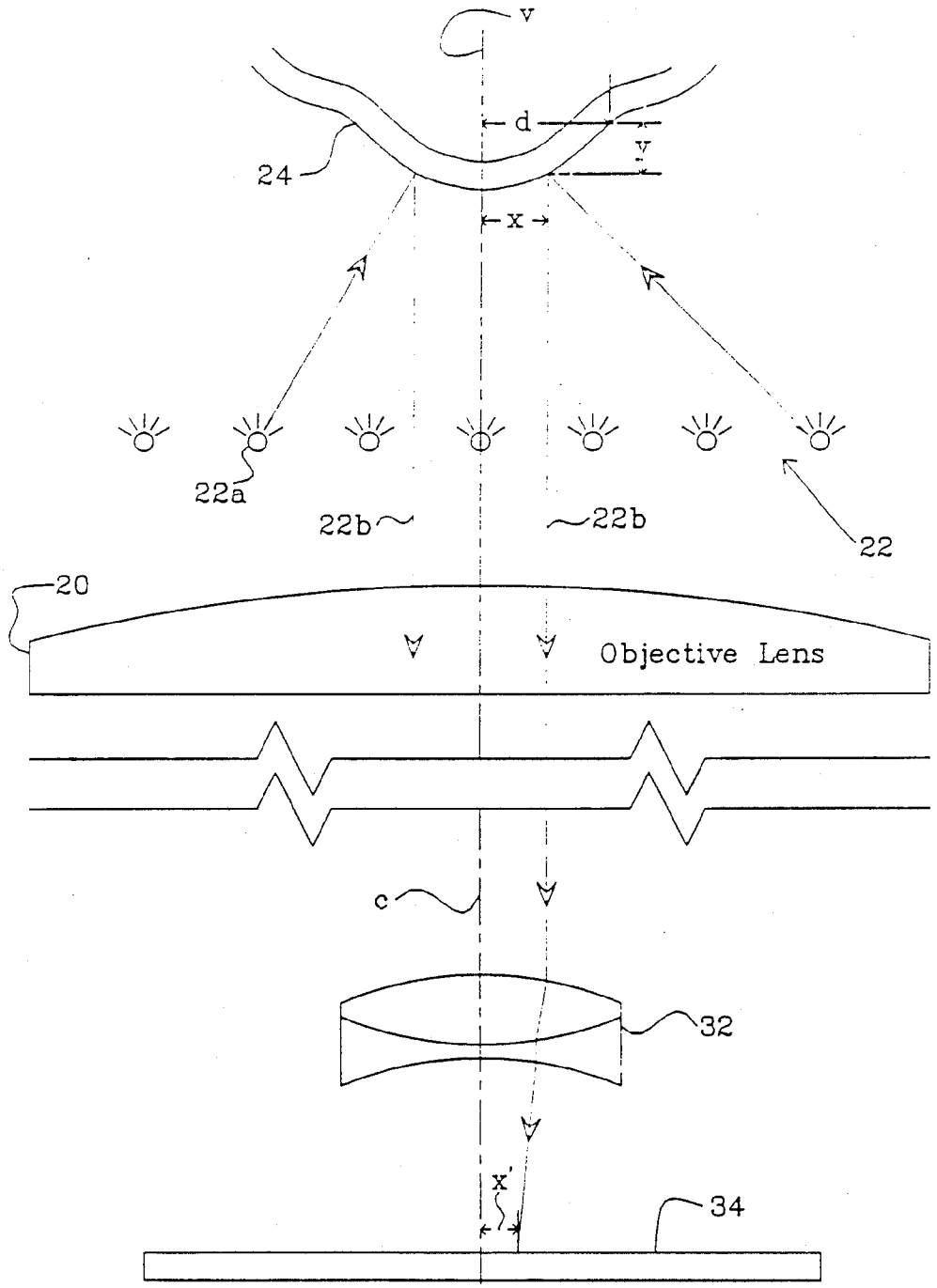



Figure 1A

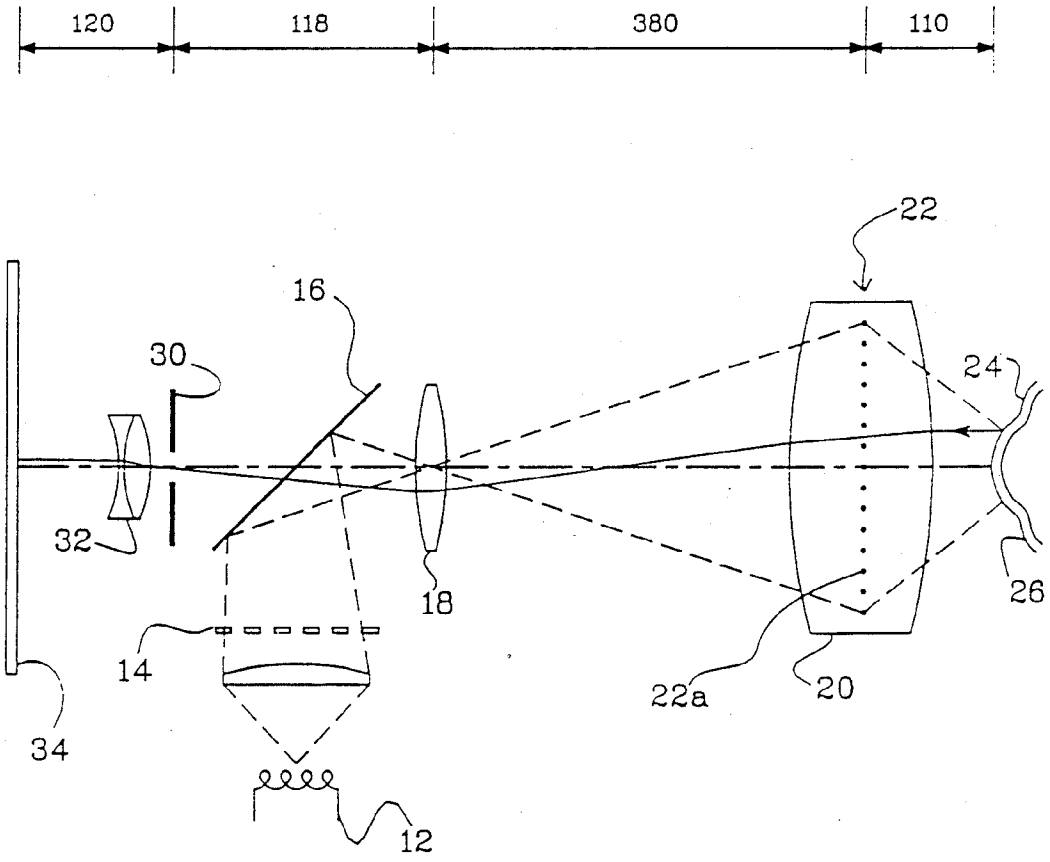



Figure 1B

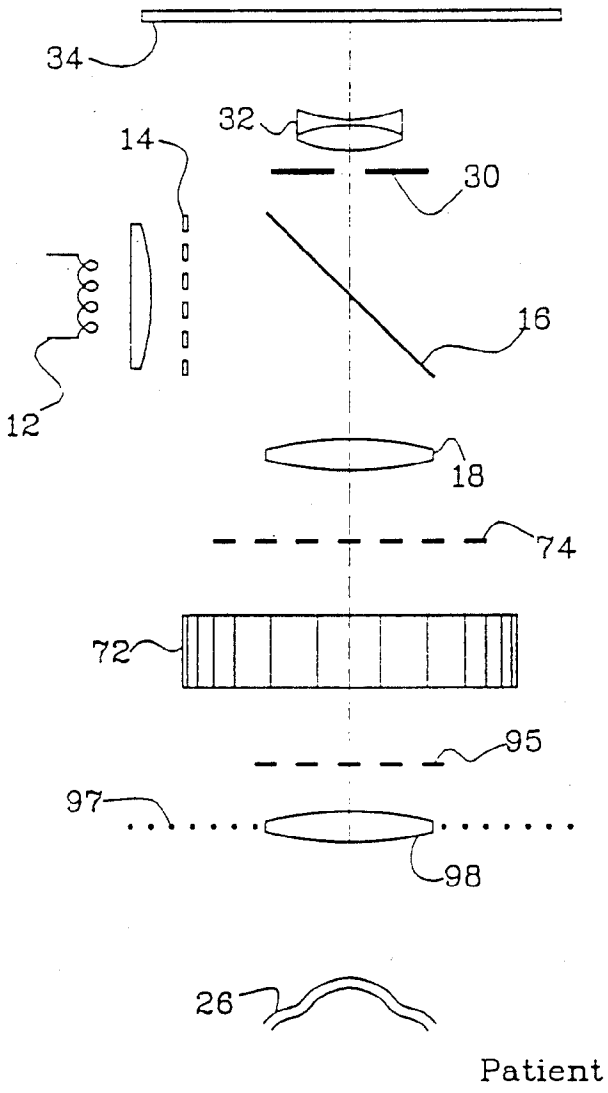
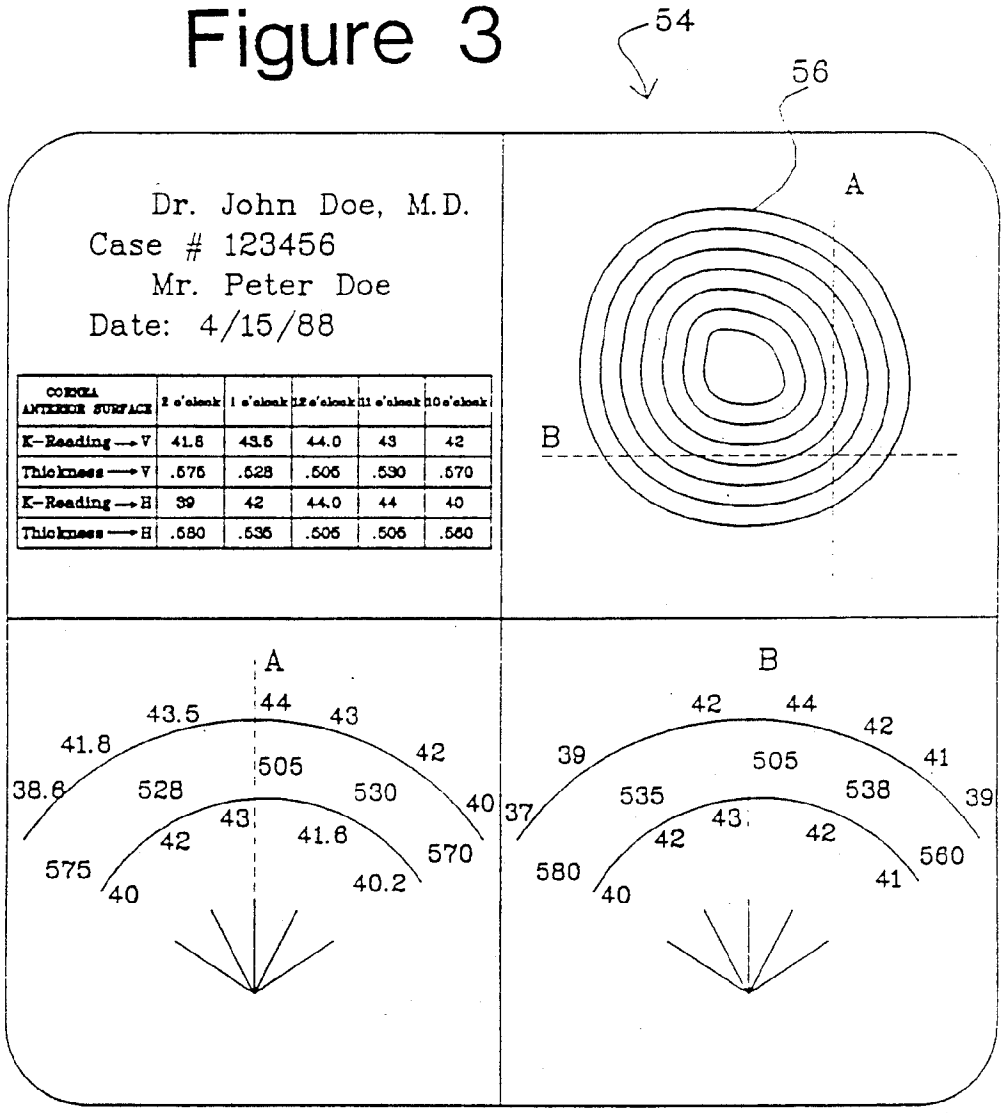



Figure 2

Figure 3




| CORNEA ANTERIOR SURFACE | 2 o'clock | 1 o'clock | 12 o'clock | 11 o'clock | 10 o'clock |
|---|---|---|---|---|---|
| K–Reading ⟶ V | 41.8 | 43.5 | 44.0 | 43 | 42 |
| Thickness ⟶ V | .575 | .528 | .505 | .530 | .570 |
| K–Reading ⟶ H | 39 | 42 | 44.0 | 41 | 40 |
| Thickness ⟶ H | .580 | .535 | .505 | .538 | .560 |

Figure 4

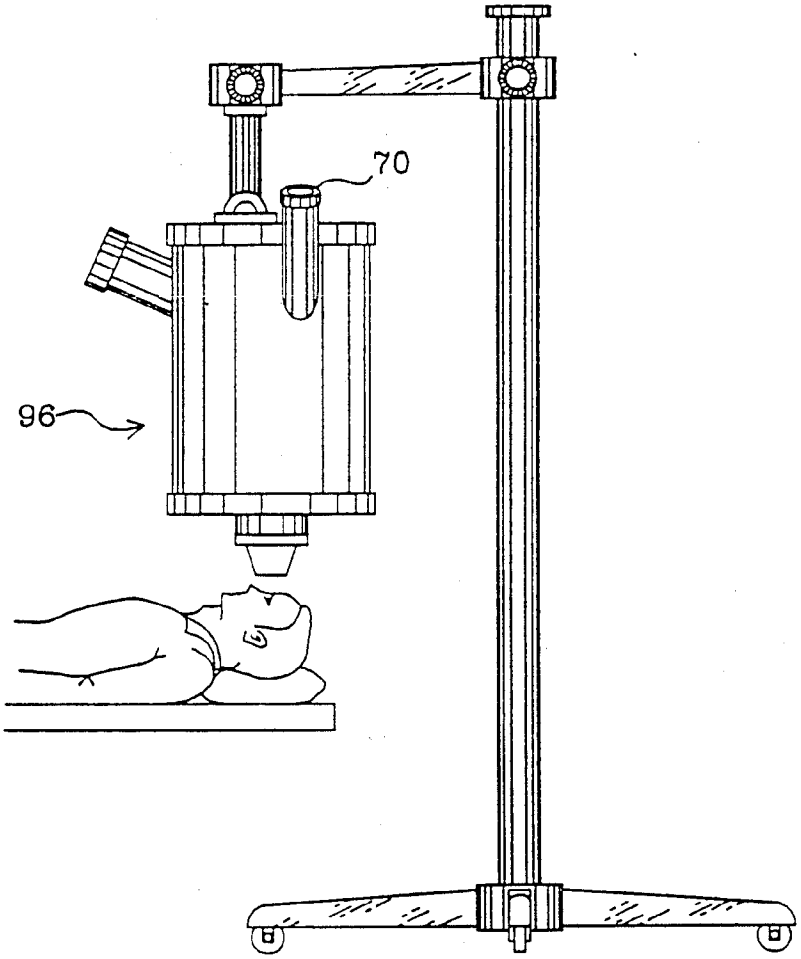



Figure 5

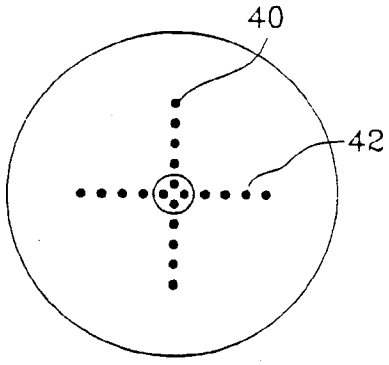



Figure 6

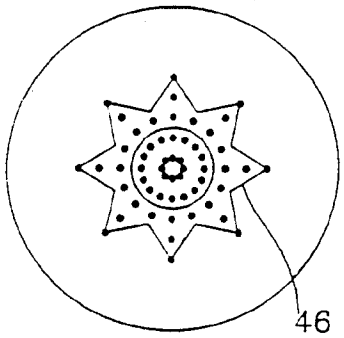



Figure 9

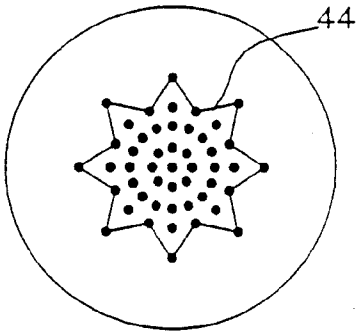



Figure 7

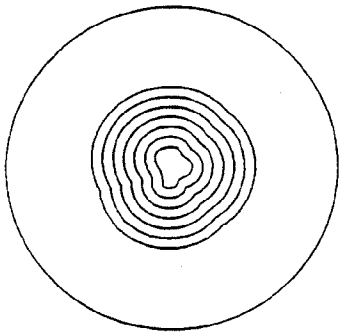


Figure 10

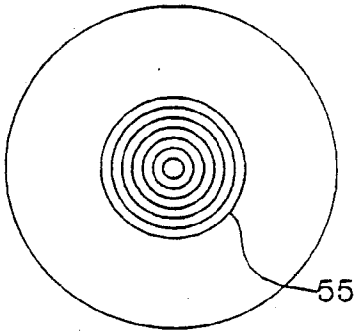



Figure 8

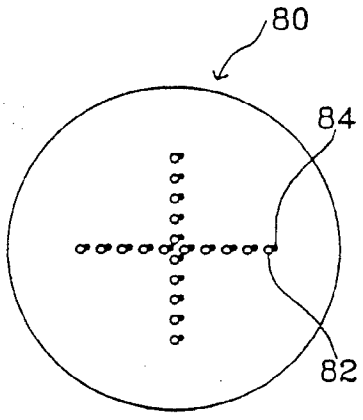



Figure 11

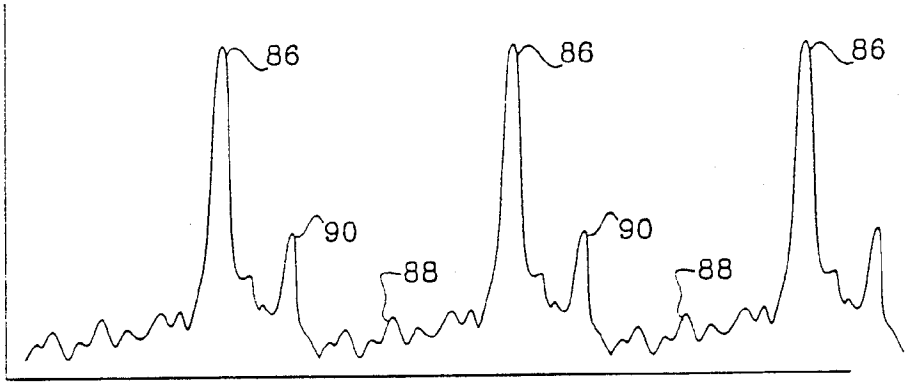



Figure 12

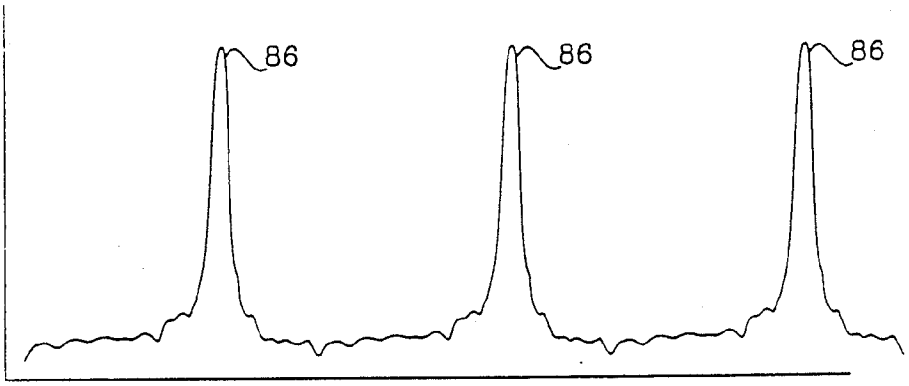



Figure 13

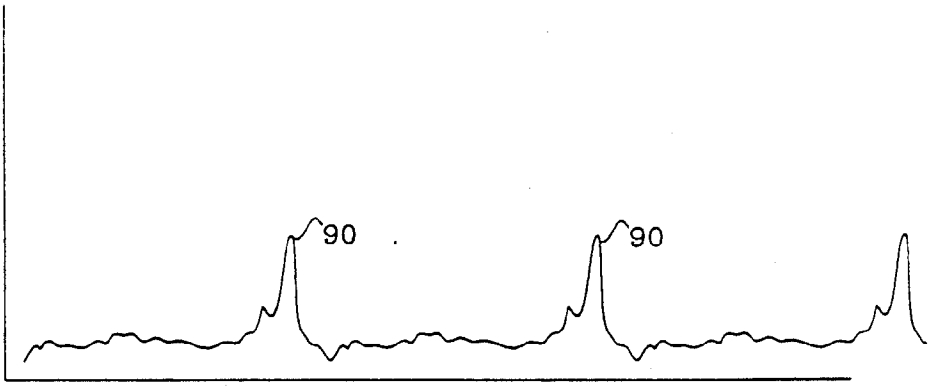



Figure 14

# OPHTHALMIC DIAGNOSTIC APPARATUS AND METHOD

## BACKGROUND OF THE INVENTION

This invention relates to ophthalmic analytical and diagnostic systems, and in particular the invention is concerned with obtaining accurate determinations of the shape of the human eye structures such as the cornea, the lens and the retina. An apparatus in accordance with the invention measures, calculates and displays the shape of selected cross sections of the cornea, for example, and is intended for use by ophthalmic surgeons as well as the eye care community at large.

One of the principal activities of the eye care specialist, which includes both ophthalmologists and optometrists, is to determine the refractive power of the eye as an optical system. Since the only major refractive index change along a light path entering the eye to impinge on the retina occurs on the first air to tear layer interface or, approximately, at the corneal anterior epithelial surface, the precise measurement of the shape of the corneal epithelial surface is the key to estimating the refractive power of a given eye.

Traditionally, the eye care specialist has been satisfied with a measurement from keratometric readings. The keratometric readings ("K-readings") correspond to the curvature of the corneal epithelial surface at the intersection of the corneal epithelial surface with the central visual axis of the eye. The K-readings are usually displayed in diopter power which is proportional to the reciprocal of the radius of curvature. The K-readings provided by keratometers correspond to the curvatures at one point on the corneal epithelial surfaces along two surface rays passing through that point. Usually, the two rays match the semi-major and semi-minor axes of the eye which are the nasal-temporal (horizontal) axis and the superior-inferior (vertical) axis of the eye.

Since the first concern of the eye care specialist is central, axial vision, the K-readings, which only provide two curvature measurements along the semi-major and semi-minor axis normal to the visual axis, represent a fair estimate of refractive power along the most critical light paths in the human eye.

To date, the eye care community has relied on the eye surface being approximated as a combination of a sphere and a cylinder, thus the reference to 20/20 as a visual standard. This approximation is exact at the intersection of the corneal anterior surface with the visual axis. The approximation is known to fail as one proceeds radially outward from the central visual axis towards the limbus, roughly the outermost edge of the cornea where the triple-point transitions between cornea, sclera, and iris tissues take place.

There are several instruments for measuring the location of the corneal anterior surface in proximity of the limbus as well as in the central region, but the display generated from these measurements usually assumes that the eye can also be approximated as a combination of spheres and cylinders. In a sense, these instruments spread the error in approximating the shape of the corneal surface from being concentrated toward the limbus to being distributed over a greater region.

Notable exceptions are instruments based on confocal microscopy that measure the actual curvatures without simplifying assumptions. However, systems based on confocal microscopy have very limited fields of view, considerably smaller than the full corneal surface. Such systems must then rely on a sequence of measurements over time which are subsequently made to piece together using either fractal techniques or some boundary matching algorithm. These paste-ups involve some form of interpolation, albeit on the boundary of the images rather than in the interior. In contrast with interpolations which are based on a sparse set of measurements, confocal techniques provide dense measurements at the expense of not having them performed simultaneously. Even though the sequential measurements can be formed very quickly, involuntary eye motion is known to occur within millisecond time scales, faster than the time required to complete data gathering using confocal techniques. This can introduce errors.

Since only a finite number of measurements of the actual location of the corneal anterior surface are possible, interpolation techniques are an intrinsic part of displaying a continuous shape based on the measured information. In most instruments still in use by the eye care community, the error in measuring the shape of the corneal anterior surface is often not in the measurement technique but in the numerical interpolation techniques utilized to prepare the display of the continuous corneal surface cross-section.

## SUMMARY OF THE INVENTION

In accordance with the present invention, a system, apparatus and method for corneal shape determination and other eye structures achieves a high degree of accuracy by real-time measurements, calculations and display.

An interpolation technique used in accordance with the invention is based upon not only measuring the location of a given set of points on the surface of the corneal tear layer, but in performing the measurement in such a manner as to be able to solve the non-linear ordinary differential equation describing the surface in real time. This technique further provides several higher order derivatives which are used in generating the continuous corneal shape.

Aside from improving the K-readings while simultaneously reducing the global error in approximating the shape of the cornea's anterior surface, the process and apparatus of this invention are designed to provide the actual corneal shape in real time. By real time is meant faster than an eye care practitioner's capacity to observe and recognize a change in corneal shape, generally on the order of one second or less.

Several instruments have claimed capability of measuring a corneal cross-section shape in real time, as defined herein. These have relied on a combination of scanning techniques, confocal microscopy techniques, and image reconstruction techniques. Namely, using confocal microscopy, a high-resolution set of measurements is taken of selected small segments of the corneal cross section and the shape of the individual crosssection pieces is analyzed and then fitted together with the other separately scanned shapes to provide an approximating shape.

Given that the human eye is constantly in motion either through voluntary or involuntary actions and that these motions do not necessarily correspond to rigid body motions, scanning techniques which later patch up different pieces of the corneal cross section may have errors introduced by eye motions and eye

distortions. The errors may be diminished by utilizing high-speed scanners, but are not removed.

The advantage of the described scanning technique is that it allows for performing a high number of measurements in a small region, thus generating very high resolution and accurate definition of the shapes in the limited area observed. To achieve the same resolution globally, that is from limbus to limbus, would require an unwieldy instrument with many light sources and measurement points. Hence speed is achieved with the scanning technique as well as high local resolution at the expense of global uncertainty due to the patchwork effort.

In one of the embodiments of the present invention, the instrument achieves comparable high local resolution without risking the uncertainty of global accuracy. This is achieved by performing a global measurement whereby the field of view of the instrument is adjustable. Thus, when global information is desired as to the shape of the cornea, one measurement rather than a sequence of scans is performed. And when a high resolution measurement is required of a particular region of the corneal surface, one of the embodiments of the invention will be able to narrow the same number of measurements into a limited field surrounding the desired corneal region.

The approach is comparable to the use of photolithography to generate pattern cuts in the semiconductor industry in the sense that a large template is progressively focused into smaller regions to generate tighter and smaller effects. In preferred embodiments of the present invention a zooming technique is used to define the field of view, and a corresponding zooming technique is used to enlarge the field to fill the display monitor.

One of the important considerations to ophthalmologists seeking to perform corneal reconstructive surgery, whether by radial keratotomy, corneal epikeratoplasty, keratimileusis, keratophakia, wide area laser ablation, or other procedures, is to accurately and reliably measure the shape of the cornea. This is important not only prior to the initiation of a surgical procedure, but also during evolution of the shape in the surgical procedure, and after the surgery, as the healing process takes over postoperatively.

One of the features of the present invention is not only to satisfy the corneal global measurement needs of pre- and postoperative surgical procedures, but to likewise provide the high resolution needed to follow the effects of surgery during the course of the procedure. For example, to determine the depth of an incision during a radial keratotomy procedure, the surgeon preselects the depth of the protruding diamond blade from the scalpel and depresses the full depth of the blade into the cornea. Surgeons are therefore relying that a uniform thrust pressure on the blade point can be maintained as the scalpel is traversed over the cornea and that the shape of the cornea does not deform during the course of the procedure. It has been often observed that the cornea deforms during the course of radial keratotomy. Thus, the surgeon is left guessing and relying on his own intuition as to the incision depth of radial cuts. The depth of the incisions becomes progressively uncertain with each successive cut because of the increased deformation.

In one of the embodiments of the invention, the surgeon is enabled to first determine the corneal shape immediately prior to commencing the procedure, then to observe the measured incision depth of each cut as it is being performed, to readjust his blade progressively as needed based upon actual incision depth measurement, and then once again to provide a global measurement of the resulting corneal shape.

Another problem addressed by the present invention again involves the utility of a measurement device during surgery. Keratometers have currently been used in surgical theaters, such as the Terry keratometer, but their efficacy has been limited not only by the extent of the information provided, but by their accuracy and by the obstruction of the patient's eye to the surgeon while the measurement is being performed. One of the reasons for this obstruction is mandated by the need to provide a field of illuminated points which the keratometer then detects as reflections from the cornea.

In order to get reflections or data points near the central visual axis, keratometers have needed to place such illumination points near the central visual axis. These illuminators are bulky and get in the way of the surgeon's access to the eye on which he is operating. In one of the embodiments of the present invention, this problem is solved by placing the illuminators a considerable distance away from the eye, folding an image of the illuminators into either a surgical microscope or other imaging apparatus via a beam splitter, and then rather than physically placing the illuminators along the visual axis, the present invention projects a real image of the illumination points at the location where the illuminators would have been required. This location is between the patient's eye and an objective lens of the instrument, or of a surgical microscope to which the instrument is attached. The real image is reflected off the surface or surfaces of interest in the eye, and reflected illumination points parallel to the optical axis of the instrument are collected and detected through the instrument. A real-time display is generated, preferably with ocular cross sections as selected by the surgeon, along with numerical topological data.

An important aspect of the present invention is that the optics of the system use the objective lens as a field lens for the pattern image and that the optics relay the Fourier plane of the objective lens, located behind the objective lens in the system, to a relayed, distant position in the instrument. This gives the opportunity and the spatial distance to fold in one or two light source patterns, between the Fourier plane of the objective lens and the relayed or transferred position of the Fourier plane. In one embodiment, as discussed below, two different light source patterns may be folded into the optical axis of the instrument, using two different beam splitters.

In this regard, two separate displays can be formed for real time review by the surgeon. One, a qualitative image showing elevation contours of the eye, derived from a series of concentric light rings; and the other from a selected pattern of discrete points of light, for quantitative analysis in producing cross sectional representations of the shape of the cornea.

The utility of the invention is not restricted to improving radial keratotomy procedures. Any surgical procedure which seeks to alter the refractive power of the eye benefits from having accurate displays showing the course and effect of the procedure. More generally, any surgical procedure which invades the eye and which in turn necessitates wound closure can be greatly benefited by the process and apparatus of the invention. Also, the instrument can be used for purely diagnostic

purposes, such as by an optometrist for fitting contact lenses.

The various embodiments described correspond to different configurations depending on the actual needs of the surgeon. The common feature is to provide high resolution wherever it is mandated while preserving computational speed and global accuracy wherever extreme resolution can be relaxed.

It is therefore among the objects of the invention to enhance ophthalmic diagnostic and surgical procedures by providing an apparatus, system and method for high speed, real time precision monitoring of the shape of the cornea, both epithelial and endothelial surfaces, and of other ocular surfaces. These and other objects, advantages and features of the invention will be apparent from the following description of preferred embodiments, considered along with the accompanying drawings.

## DESCRIPTION OF THE DRAWINGS

FIG. **1** is a schematic diagram showing the layout of an ocular diagnostic system and apparatus in accordance with a preferred embodiment of the present invention. The figure shows a series of optical elements which can be incorporated in this embodiment of the invention.

FIG. **1A** is a schematic diagram showing in greater detail certain portions of the system of FIG. **1**.

FIG. **1B** is a view similar to FIG. **1**, but showing a slightly modified form of the system, and with an example of distances and other optical values which can be used in the system.

FIG. **2** is a schematic diagram showing portions of the apparatus of FIG. **1** as they can be incorporated into a surgical microscope, preferably with a simple auxiliary camera mount connection (for example, a C-mount connection).

FIG. **3** .is a schematic view showing an example of a video display which can be presented as a result of the information gathered by the instrument of the invention.

FIG. **4** is a view showing an example of quantitative information which can be displayed to the user of the instrument.

FIG. **5** is a view showing a conventional surgical microscope with which the embodiment of FIG. **2** can be used.

FIG. **6** is a schematic view showing a simple pattern of rectilinear sequences of point light sources which can be used for measuring the cornea in accordance with one embodiment of the invention.

FIG. **7** is a schematic view similar to FIG. **6**, showing another pattern which can be used for measuring the cornea.

FIG. **8** is another schematic view showing concentric circles of light which can be projected on the cornea simultaneously with the pattern shown in FIG. **6** or FIG. **7**, strictly for qualitative information for the surgeon, for comparing against the measurements determined from the quantitative measurements obtained via the pattern of FIG. **6** or FIG. **7**.

FIG. **9** is a schematic view showing a distortion of the pattern shown in FIG. **7**, as an example of what may be read and analyzed by the apparatus of the invention for determining cornea shape.

FIG. **10** is another schematic view, showing a reflected pattern as produced by the projection shown in FIG. **8**, and the distortion of the reflected pattern.

FIG. **11** is a schematic view showing a reflected pattern of discrete point light sources which might be received from the measurement projection shown in FIG. **6**, and also showing an example of a secondary reflection which is taken from the endothelial surface, i.e. the back surface of the cornea.

FIGS. **12**, **13** and **14** are schematic graph plottings showing examples of light intensity versus distance across the cornea, for a reflected projection such as shown in FIG. **11**, and indicating analysis of these reflections to obtain information as to both the front surface and the back surface of the cornea, i.e. the epithelial and endothelial surfaces.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the drawings, FIG. **1** shows in schematic representation a system of optical elements in accordance with the invention for use in carrying out ophthalmic diagnosis and analysis.

The system, generally identified by the reference number **10**, includes an illuminator or light source **12**, a pattern plate or disk **14** having a pattern of holes cut in the plate for producing a desired pattern of discrete light sources, a non-distorting beam splitter **16**, a lens **18** which projects an image of target **14** onto an image plane at **22**. This image plane **22** is close to or coincidental with the system of objective lens **20**. The purpose of placing the image at this location **22** is to have the objective lens **20** serve as a field lens, that is bending the rays of light that form the image towards the patient's cornea **24**.

As indicated in FIG. **1**, the focused image **22** of the pattern is a real image, formed at some plane at or near the lens **20** and between the lens and the patient. The real image preferably is in the lens **20**, but it can be very closely in front of the lens (i.e. a few millimeters in front). In this real image, each point source of light **22***a* projects a cone of light toward the patient. Thus, each point source **22***a* in the real image makes an infinite number of specular reflections off the front surface of the cornea **24** of a patient's eye **26**. As explained below, the F-number of the final lens **20** determines the maximum area of the cornea that can be measured. The objective lens serves as a field lens, and the patient's cornea must be at the focal length of the lens **20**. This assures that the light reflected off the eye parallel to the optical axis of the instrument is then brought to a point behind the lens **20** at the focal distance of the lens **20**. This enables the return light to be apertured down as discussed below, to select only those rays which were paraxial off the eye. This enables the system to localize a detected point to a point on the cornea from which that ray was reflected. If the objective lens **20** were not situated to serve as a field lens, outermost points of light in the pattern would not reflect off the cornea. As a field lens, the lens **20** efficiently bends the outer points of light toward the eye.

It is preferred that the focal length of the lens be great enough to provide an unobstructed, comfortable distance from the instrument to the patient and adequate working room for the surgeon, for surgical applications.

The F-number of the objective lens **20** is most important in its function as a field lens as it will determine the maximum angle from the optical axis at which a ray can be reflected from the cornea parallel to the axis. If for example a commercially available F/2 lens is used, then the region of coverage will be about 3 mm diameter on

the cornea. A lens with a smaller F-number will cover a proportionally larger region on the cornea.

As indicated in FIG. 1A, each of the real-image point light sources **22***a* makes at least one reflection **22***b* which will be parallel to the central axis of the objective lens **20**, with all axial reflected rays **22***b* parallel as shown in the drawing. For each point light source **22***a*, the reflected axial ray **22***b* will be unique unless the corneal surface has extremely strong local imperfections or distortions in the corneal curvature, which could theoretically cause more than one reflected axial ray **22***b* to occur, from spaced locations on the cornea.

Other rays of light reflected off the cornea will reach and pass through the lens **20**, but as will be seen below, only those returning reflected rays which are very nearly parallel are passed through the system for analysis. Those are the rays and points which will supply data points to be compared with the original pattern as projected through the plate **14** to supply data which can be solved to determine the shape of the cornea.

As shown in the overall schematic view of FIG. **1**, the returning reflected rays pass back through the lens **20**, then through the lens **18** and the beam splitter **16**, an aperture or spatial filter **30** and a further lens **32**, ultimately to be focused on a detector or camera plane **34**.

The curvature of the cornea **24** forms a virtual image **93** of the target image **22**. In the article "Suggested New Methods for Photokeratoscopy, a Comparison for Their Validities, Part I", by S. G. El Hage, *American Journal of Optometry and Archives of American Academy of Optometry*, November 1971, El Hage pointed out that an aperture or spatial filter at the back focal plane or Fourier plane of the objective lens **20** will only pass rays parallel to the axis thus localizing those rays from a given point of the virtual image **93** to those that are reflected from a specific point on the cornea **24**. In this embodiment, it is desired to have space behind the objective lens **20**, the lens **18** is used to relay the Fourier plane of the lens **20** to the aperture **30**. The aperture **30**, being in an image of the Fourier plane, will likewise select only those rays reflected from the cornea **24** parallel to the axis.

The rear lens **32** of the system focuses a distorted image of the virtual image **93** of point light reflections on the detector or camera plane **34**.

As shown schematically in FIG. **1A**, the camera plane **34** has a central axis c which lies on the optical axis of the system, including the objective lens. Ideally this axis is placed as closely as possible to the center of the cornea or visual axis v. If these axes are significantly displaced, then much of the light reflecting off the cornea will not be returned through the system. This discussion assumes the axes coincide, but adequate information can be obtained over small deviations (e.g. one millimeter). If a reflected, returned point lies on the center axis c of the camera plane **34**, then that ray emanated from the visual axis v of the cornea, at least as respects one orthogonal direction on the cornea and on the camera plane **34**, which is shown as the left-right direction in the plane of the paper in FIG. **1A**.

Likewise, if a particular point of light is focused on to the camera plane or detector face **34** at a distance x' from the center axis c, that distance corresponds to, and is linearly proportional to, a distance x of the reflecting point on the cornea for that ray **22***b* as measured from the visual axis v. If a depth distance y is determined, measured from an arbitrarily chosen datum d to the point of reflection on the cornea, and a series of such x and y can be determined, then a differential equation can be solved to define y as a function of x, giving the curvature of the cornea in this direction or along the subject axis, i.e. in the plane of FIG. **1A**. Similarly, measurements and calculations can be made along an orthogonal axis on the cornea (e.g. the nasal-temporal and superior-inferior axes can be used), giving as much information regarding the cornea's shape as is normally needed for any diagnostic or surgical procedure.

The y distance indicated in FIG. **1A** can be derived through information regarding the degree of distortion of the reflected point light pattern, and the spatial relationship among the points of light, as compared to the pattern as originally projected and as arranged in the real image **22**. Thus, considering the parallel ray **22***b* in FIG. **1A**, which is shown as emanating from the real image point light source **22***a* on the right in FIG. **1A**, if the cornea curvature is less steep at the point of reflection, i.e. at a shallower angle with respect to a tangent to the cornea at the visual axis, then the parallel ray **22***b* would have originated from a different real image point source, one farther to the left in the pattern. The right-end point source **22***a* would have created a parallel reflection only from another point on the cornea, farther to the right as viewed in FIG. **1A**. Each of the reflected points as detected at the camera plane **34** can be identified electronically, essentially by counting points in the array.

FIG. **6** shows one example of a projected light pattern which can be used in the system and method of the invention. In this simple pattern, a vertical rectilinear array **40** is crossed orthogonally with a horizontal rectilinear array **42**, with the intersection point corresponding to the visual axis of the eye. This is the simple pattern assumed with reference to FIG. **1A**.

A more complex pattern **44** of points is shown in FIG. **7**. This pattern, shown as an asterisk-like pattern of linear arrays of points, gives data from many more points on the cornea. It may define an outline of a five-pointed star or any similar type of pattern, but preferably it has some means of identifying its rotational orientation. It may have an outline of a star with an odd number of points, so that the asymmetry can help identify the detected, reflected points by correlating them with the originally projected pattern **44**. FIG. **9** shows an example of a reflected pattern **46** which might result from the pattern **44** shown in FIG. **7**, as reflected from a cornea with some degree of distortion.

FIG. **1** schematically indicates that the detector or camera plane **34** is connected to a microprocessor **50**. The microprocessor may be connected to a display device, such as a CRT monitor **52** as indicated. Data gathered from the system as described is received by the microprocessor **50** and analyzed. Each detected point is correlated with the location of the particular point in the source pattern from which it emanated. The x value is determined for each point, i.e. the distance from the optical axis v from which the point was reflected off the cornea. This is determined by direct proportioning, from the known magnification of the system. Each reflected point has an x value which is the distance from the optical axis of the system. Each linear array of points in the image must be separately analyzed and fitted to the mathematical approximation. If the complex pattern **44** shown in FIG. **7** is used, formed of an asterisk-like array, the analysis and computation are made along each line of the pattern.

By the method and system of the invention, the mathematical shape of the cornea is determined by assuming an analytical approximation to the surface shape. The analytical approximation is then substituted into a differential equation and some type of appropriate fit is performed to determine the coefficients that satisfy the differential equation. In a preferred embodiment of the invention, a nonlinear least squares fit is performed.

These operations are performed in the microprocessor **50**. The processor has programming to review a great number of x values as determined on the detector **34**, substituting all of these values into the differential equation and arriving at a formula for y and as a function of x.

A differential equation suitable for this purpose is

$$dy/dx = -\left(\frac{a(y)-x}{b(y)-y}\right) \pm \left[\left(\frac{a(y)-x}{b(y)-y}\right)^2 + 1\right]^{\frac{1}{2}}.$$

where y is the depth of the reflection site away from a datum plane (such as the datum plane d shown in FIG. 1A), x is the distance from the visual axis, and a and b are coordinates representing the location of the real image of the illumination point in space. a is a distance of the particular illumination point **22***a* (see FIG. 1A) from the visual axis and b is the depth of that illumination point out from the datum plane d.

The differential equation used in this process is not new. It is a general equation which can be used to represent the shape of any surface, and is described in the article "Suggested New Methods for Photokeratoscopy, a Comparison for Their Validities, Part I", by S. G. El Hage, *American Journal of Optometry and Archives of American Academy of Optometry*, November 1971, page 897. In the article, El Hage discusses various uses of this general equation for solving the shape of the corneal surface. Also, he relates the corneal surface shape to one of the keratoscope rings in photokeratoscopy. Thus, this derivation in itself does not form a part of the present invention, but is hereby incorporated in this application by reference as illustrating that such derivation is known in the art.

At page 909, El Hage shows an optical arrangement for projecting an image onto a cornea and for detecting reflected light from the cornea. His source is analogous to the real image in the present invention, and El Hage had a number of optical elements between the source and the eye, including a beam splitter between the objective lens and the eye.

Returning to FIG. **1**, the illuminating light source **12** may be a visible light source, in preferred embodiments of the invention wherein the system is not combined with a coaxial surgical laser. For example, an incandescent lamp can be used. The pattern plate or target **14** may be laser or photolithographically cut, with hole sizes on the range of about thirty microns. The beam splitter **16** may be a simple nondistorting plate glass beam splitter, with a surface coating of about 50% reflectivity.

In one specific embodiment of the invention, particular lenses and lens relationships may be selected as indicated in FIG. **1B**. In FIG. **1B** the distances between lenses, focal lengths and diameters of the various lenses are given for this specific embodiment. Other relationships and distances are also given, including the diameter of the aperture or spatial filter **30**. The system of FIG. **1B** shows a single light source **12** projecting a pattern and being folded into the optical axis of the instrument.

The detector or camera plane may comprise a high density photodetector array, for example.

As indicated schematically in FIG. **1**, the microprocessor **50** is connected to a display monitor **52**. An example of the type of display that can be presented to the physician in real time is indicated in FIG. **3** by the reference number **54**. In the upper left quadrant of the screen, patient identifying data is given, along with K-readings and thickness readings. A more detailed example of this information is shown in FIG. **4**.

The lower left and lower right quadrants of the display **54** show examples of depth references of the epithelial and endothelial cornea surfaces at cutting planes A and B shown in the plan view of the upper right quadrant. The location of these cutting planes is preferably selectable by the physician, via inputs to the microprocessor **50** (not shown).

The distorted image **56** shown in the upper right quadrant of FIG. **3** is derived from a second projection which is preferably included in preferred embodiments of the invention. As illustrated in FIG. **1**, a second projection may be folded onto the axis of the lens system via a second beam splitter **60**, which reflects light from an illuminator light source **60** to a pattern or mask **64**. The mask **64** has a plurality of concentric circle cuts so as to project a real image of the concentric circles in front of the cornea as is done with the pattern **22** of point light sources.

FIG. **8** shows schematically a series of concentric circles in a pattern **55** which can be projected via the pattern plate **64**. The detector **34**, which may be a pixel array of very high density, can receive and detect both reflected images simultaneously. The concentric ring pattern can be discerned from the point source pattern by the contiguity of each ring. The software employed by the microprocessor **50** can sample each pixel receiving light and determine whether any immediately adjacent pixel is also receiving light. If so, the contiguity of a ring is indicated. In contrast, the patterns of point light sources such as shown in FIGS. **6** and **7** will not display appreciable contiguity. Thus, the microprocessor **50** can separate these images and analyze each separately. Alternatively, in a separate embodiment of the system and additional camera detector can be placed together with an additional beam splitter to separate the image of the continuous rings from the image of the discrete point sources.

As in a conventional corneoscope or in using a Placido ring, the concentric light rings produce a reflection off the cornea which is distorted in a way corresponding to distortions on the corneal surface. This can result, for example, in a pattern of distortion **56** such as shown in FIG. **3**.

FIG. **1** shows that, with two different light patterns folded into the system, onto the axis of the lenses **18** and **20**, polarizers **66** and **68** should be used to establish opposite polarity for the two different images being projected.

A polarizer used as an analyzer **94** may be rotated to select either of the projected images.

FIG. **2** is a schematic representation of an alternate embodiment showing some of the same elements present in the embodiment of FIG. **1**, but in an arrangement for connection directly with a surgical microscope. Surgical microscopes, such as those made by Weck,

Nikon, Topcon, Zeiss, Nidek, or Wild, usually include a standard auxiliary bayonet mount or screw attachment for a camera. FIG. **5** shows a typical standard surgical microscope. An auxiliary mount **70** (e.g. a C-mount) is shown in FIG. **5** and schematically indicated in FIG. **2** as coupling the system embodying the elements **12**, **4**, **34**, **32**, **18** and **16** to a fitting or optical tube **72** on the surgical microscope. Generally the surgical microscope will have optics to produce an image at an image plane **74** which is a standard distance from the auxiliary mount on the fitting **72**, for coupling a standard **35** millimeter video camera to the surgical microscope. Thus, in this embodiment of the present invention, the objective lens **20** is eliminated and replaced by the objective lens **96** of the surgical microscope. The focal length of the lens **18** is adjusted to appropriately relay the Fourier plane **95** of the surgical microscope lens **96** to the aperture plane **30**. In almost all other respects this embodiment is similar to the previously described embodiment. One possible exception is that if the F-number of the surgical microscope objective lens **96** is not sufficiently low to give the desired area of coverage on the cornea, then additional point sources **97** of light at multiple locations will be necessary outside the objective. These additional sources may be created with an illuminated pinhole mask or optical fibers.

FIGS. **11** through **14** illustrate an aspect of the system of the invention which enables both the epithelial corneal surface and the endothelial corneal surface to be detected and displayed in real time simultaneously. FIG. **11** shows an example of a reflected pattern **80** which might occur at the detector **34** from the simple pattern shown in FIG. **6** comprising a pair of orthogonal linear arrays of light points. As indicated in FIG. **11** each detected point **82** which is not on the optical axis will have a secondary reflection **84**, of much lower intensity emanating from the back surface or endothelial surface of the cornea. The detected array might produce, for example, an intensity versus distance curve such as shown in FIG. **12**. The long spikes **86** of light intensity represent the reflection of the discrete point light sources from the anterior, or front, surface of the cornea, with some degree of noise **88** occurring between the spikes. A secondary spike or cluster **90** of light intensity which is discernibly higher than the noise **88** occurs adjacent to each high intensity spike **86**. This represents the lower-intensity reflection of the light points off the endothelial cornea surface. The plotting shown in FIG. **12** can easily be sampled or filtered to identify and separate the high intensity spikes **86** from the low intensity spikes **90**. As can be appreciated by those skilled in this art, the programming in the computer can first determine the signal contribution from those spikes which achieve amplitudes above a predetermined threshold and then subtracting the contribution to the signal which correspond the high intensity spikes **86** to obtain a signal which contains only the low intensity spikes **90** and the noise **88**. The process of identifying spike location for the high intensity spikes **86** is now repeated for the low intensity spikes **90**, but with a lower threshold. In some embodiments of the invention, it may prove efficacious to electronically amplify the signal from which high intensity spikes **86** have previously been deducted in order to facilitate the threshold differentiation between low intensity spikes **90** from the noise **88**. It is important to note that this selection process is facilitated by the observation that the specific amplitude of the spikes **86** and **90** are not as important as their actual location.

FIG. **13** and **14** show separate plottings of intensity versus distance for the front surface, anterior reflection and the rear surface, posterior reflection.

Once the primary and secondary reflections are known and located as in FIGS. **13** and **14**, the shapes and elevation points of both the epithelial and endothelial surface can be calculated by the approximation method described previously, and two sets of data can thus be presented to the physician. Similarly, the cross sections and appropriate values can be represented in the lower two quadrants of the display as illustrated in FIG. **3**.

It should be understood that in the drawings and the description herein, as well as in the claims, references to "up", "down", "lower", "upper", "left" or "right" are intended only for convenience in referring to the embodiments as represented in the drawings, and not as limiting any possible orientations of the instrument or components. The drawing figures are not to scale. Further, the term "objective lens" as used herein and in the claims and drawing figures is intended to refer to either an objective lens specific to the instrument or an objective or final focusing lens of a surgical microscope, if the instrument is used as part of a surgical microscope.

The above described preferred embodiments are intended to illustrate the principles of the invention, but not to limit its scope. Other embodiments and variations to these preferred embodiments will be apparent to those skilled in the art and may be made without departing from the spirit and scope of the invention as defined in the following claims.

We claim:

**1**. An ophthalmic diagnostic instrument for determining the shape of the cornea of a patient's eye, comprising,

an objective lens as an optical element of the instrument, on an optical axis of the instrument,

means for projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located between the interior of the objective lens and the patient's eye, the real image including point light sources in positions which traverse substantially directly across the optical axis,

means for expanding the region of coverage on the cornea by using the objective lens as a field lens for the pattern image,

means for selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and for detecting a reflected position of substantially each point light source, as reflected from the cornea, including means for relaying the Fourier plane of the objective lens to a relayed position in the instrument, with aperture means positioned at said relayed position for limiting the collected light to that which is reflected paraxially off the cornea, whereby the aperture means is a spaced distance from the objective lens,

means for analyzing the returned, collected pattern image and for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected, and

means for deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image,

whereby the real image point sources, extending substantially across the optical axis, enable enhanced measurement of the central optical zone about the visual axis of the cornea.

2. Apparatus according to claim 1, wherein the pattern of discrete separated point light sources includes more than one rectilinear sequence.

3. Apparatus according to claim 1, wherein the pattern of discrete separated point light sources includes an asymmetrical shape having a plurality of lines of point light sources.

4. Apparatus according to claim 3, wherein the asymmetrical shape comprises generally a star with an odd number of points.

5. Apparatus according to claim 1, wherein the means for expanding the region of coverage comprises means for placing the source image in the objective lens, thus using it as a field lens.

6. Apparatus according to claim 1, in combination with a surgical microscope having a standard auxiliary camera mount, and the ophthalmic diagnostic instrument being connected to the surgical microscope via the auxiliary camera mount, with an objective lens of the surgical microscope serving as the objective lens of the ophthalmic diagnostic instrument.

7. Apparatus according to claim 1, including an illuminating light source, a pattern plate or mask positioned for projecting light from the light source through the plate, beam splitter means in the path of projected light from the pattern plate for reflecting and folding the projected pattern into a path coaxial with the optical axis of the instrument, with the beam splitter means located in said spaced distance between the aperture means and the objective lens, and optical means between the beam splitter means and the position of the patient's eye for focusing the projected pattern into the real image in front of the patient's eye.

8. Apparatus according to claim 7, further including means for receiving a reflected pattern from the cornea back through the optical means and through the beam splitter means, detector means on the opposite side of the beam splitter means from said optical elements, further optical means for focusing the return reflected and distorted pattern onto the detector means, and said aperture being positioned in a Fourier plane of the returning reflected pattern to eliminate substantially all light from the detector means except light reflected off the cornea as parallel to the optical axis of the instrument, whereby the spatial orientation of the pattern detected on the detector may be compared to the originally transmitted pattern for determination of the corneal shape through analysis of the positions of reflected points of the pattern.

9. Apparatus according to claim 1, wherein the ophthalmic diagnostic instrument includes means for folding the pattern of discrete separated point light sources onto the optical axis of the instrument, toward the cornea, with the means for projecting the pattern including a source of the pattern off-axis from the optical axis and from the path of the returned, distorted pattern image.

10. An ophthalmic diagnostic instrument for determining the shape of the cornea, comprising,

an objective lens as an optical element of the instrument, on an optical axis of the instrument,

means for projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located between the interior of the objective lens and the eye,

means for expanding the region of coverage on the cornea by using the objective lens as a field lens for the pattern image,

means for selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and for detecting a reflected position of substantially each point light source, as reflected from the cornea, including means for relaying the Fourier plane of the objective lens to a relayed position in the instrument, with aperture means positioned at said relayed position for limiting the collected light to that which is reflected paraxially off the cornea, whereby the aperture means is a spaced distance from the objective lens,

means for analyzing the returned, collected pattern image and for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected,

means for deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image, and

means for projecting a second light pattern comprising concentric circles toward the cornea simultaneously with said pattern of discrete separated point light sources, and means for separately analyzing distorted reflected light from the cornea relating to the concentric circles and for providing separate, qualitative information which can be compared with the corneal surface shape derived via the pattern of discrete separated point light sources.

11. Apparatus according to claim 10, wherein said means for projecting a second light pattern includes a second light pattern illuminating source, a second pattern plate or mask, second beam splitter means positioned along the axis of the instrument and in position to fold a projected pattern from the second pattern mask onto the optical axis of the instrument, and a first polarizer in the path of the projected pattern of discrete separated point light sources and a second polarizer in the path of the second light pattern, with opposite polarity established by the orientation of the two polarizers such that the two projected patterns are projected to real image locations with opposite polarity, and such that their reflections are more easily separable with a polarizing analyzer at the detector means.

12. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the back or endothelial surface of the cornea.

13. Apparatus according to claim 12, wherein said means for separately analyzing includes filtering means for electronically separating returned light points on the detector means occurring from the front surface of the cornea from those occurring from the back surface of the cornea, by separating different ranges of amplitude of the detected light.

14. Apparatus according to claim 13, wherein said means for deriving mathematically includes computer means for determining the endothelial corneal surface shape from the locations on the detector means of the

detected light points reflected from the endothelial surface.

15. Apparatus according to claim 1, wherein the means for projecting a pattern includes an illuminating light source and a plate with a laser-cut or photolithographically produced pattern of discrete holes to form the discrete separated point light sources.

16. An ophthalmic diagnostic instrument for determining the shape of the cornea, comprising,

an objective lens as an optical element of the instrument, on an optical axis of the instrument,

means for projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located between the interior of the objective lens and the eye,

means for expanding the region of coverage on the cornea by using the objective lens as a field lens for the pattern image,

means for selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and for detecting a reflected position of substantially each point light source, as reflected from the cornea, including means for relaying the Fourier plane of the objective lens to a relayed position in the instrument, with aperture means positioned at said relayed position for limiting the collected light to that which is reflected paraxially off the cornea, whereby the aperture means is a spaced distance from the objective lens,

means for analyzing the returned, collected pattern image and for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected,

means for deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image, including means for utilizing the detected spatial orientation of the reflected point light sources to determine the corneal shape along a selected cutting plane on the eye in accordance with the general formula

$$dy/dx = -\left(\frac{a(y)-x}{b(y)-y}\right) \pm \left[\left(\frac{a(y)-x}{b(y)-y}\right)^2 + 1\right]^{\frac{1}{2}},$$

whereby y is depth of a cornea reflection point from a datum plane, x is distance from the optical axis of the instrument, and (a, b) are the coordinates of the real image of an illumination point space.

17. A method for determining the shape of the cornea of an eye, comprising,

projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located in front of the eye, the real image including point light sources in positions which traverse substantially directly across the visual axis of the eye,

selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and detecting a reflected position of substantially each point light source, as reflected from the cornea,

analyzing the returned, collected pattern image for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected, and

deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image. An ophthalmic diagnostic instrument for determining the shape of the cornea, comprising,

an objective lens as an optical element of the instrument, on an optical axis of the instrument,

means for projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located between the interior of the objective lens and the eye,

means for expanding the region of coverage on the cornea by using the objective lens as a field lens for the pattern image,

means for selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and for detecting a reflected position of substantially each point light source, as reflected from the cornea, including means for relaying the Fourier plane of the objective lens to a relayed position in the instrument, with aperture means positioned at said relayed position for limiting the collected light to that which is reflected paraxially off the cornea, whereby the aperture means is a spaced distance from the objective lens,

means for analyzing the returned, collected pattern image and for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected,

means for deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image, and

whereby the real image point sources, extending substantially across the optical axis, enable enhanced measurement of the central optical zone about the visual axis of the cornea.

18. The method of claim 17, wherein the pattern of discrete separated point light sources comprises a generally cruciform shaped pattern with crossing rectilinear rays of point light sources as an intersection point lying on the optical axis of the instrument.

19. The method of claim 17, wherein the pattern of discrete separated point light sources comprises a generally asterisk shaped pattern with an intersection point at the optical axis of the instrument, and including means associated with the pattern for establishing a readily identifiable rotational orientation of the pattern.

20. The method of claim 17, including using an off-axis illuminating light source and projecting light through a pattern mask and then reflecting the projected pattern off a beam splitter to fold the projected pattern into a path coaxial with the optical axis of the instrument, focusing the projected pattern reflected off the beam splitter to form the real image in front of the patient's eye, receiving a reflected pattern from the cornea back through the beam splitter and focusing the reflected pattern onto the detector, and including passing the returning reflected light pattern through an aperture en route to the detector to eliminate all light reflected off the cornea except that which is parallel to the optical axis of the instrument, whereby the spatial orientation of the pattern detected on the detector may

be compared to the originally transmitted pattern for determination of the corneal shape through analysis of the positions of reflected points of the pattern.

**21**. A method for determining the shape of the cornea of an eye, comprising,

projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located in front of the eye,

selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and detecting a reflected position of substantially each point light source, as reflected from the cornea,

analyzing the returned, collected pattern image for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected,

deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image, and

projecting a second light pattern comprising concentric circles toward the cornea simultaneously with the pattern of discrete separated point light sources, and separately analyzing reflected light from the cornea relating to concentric circles and providing separate, qualitative information which can be compared with the corneal surface shape derived via the pattern of discrete separated point light sources.

**22**. The method of claim **17**, further including separately analyzing a secondary returned, reflected pattern image as reflected from the back or endothelial surface of the cornea.

**23**. The method of claim **22**, wherein the step of separately analyzing includes electronically separating returned light point on the detector means occurring from the front surface of the cornea from those occurring from the back surface of the cornea, by separating different ranges of amplitude of the detected light on the detector.

**24**. The method of claim **23**, including mathematically deriving the endothelial corneal surface shape on a computer, from the locations on the detector of the detected light points reflected from the endothelial surface.

**25**. A method for determining the shape of the cornea of an eye, comprising,

projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located in front of the eye,

selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and detecting a reflected position of substantially each point light source, as reflected from the cornea,

analyzing the returned, collected pattern image for comparing it to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected, and

deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image,

including utilizing the detected spatial orientation of the reflected point light sources to determine the corneal shape along a selected cutting plane on the eye in accordance with the general formula

$$dy/dx = -\left(\frac{a(y)-x}{b(y)-y}\right) \pm \left[\left(\frac{a(y)-x}{b(y)-y}\right)^2 + 1\right]^{\frac{1}{2}},$$

where y is depth of a cornea reflection point from a datum plane, x is distance from the optical axis of the instrument, and (a, b) is the coordinate location of the real image of an illumination point in space.

**26**. The method of claim **17**, further including electronically producing cross sectional images of the eye at selected cutting planes using information derived from the mathematical derivation of the corneal surface shape.

**27**. The method of claim **17**, wherein the pattern of discrete separated point light sources is projected through the objective lens of a surgical microscope, and including collecting the returned pattern image through the objective lens of the surgical microscope.

**28**. The method of claim **21**, further including producing and displaying an image showing the distortion of the projected concentric circles.

**29**. The method of claim **17**, wherein a front objective lens comprises an element closest to the eye of the patient, and including spacing the front objective lens at least about **110** mm away from the eye.

**30**. The method according to claim **17**, including forming the real image of the pattern of point light sources substantially in the objective lens, so as to use the objective lens as a field lens, enabling a larger field of view of the cornea.

* * * * *

# APPENDIX C

US005170193A

# United States Patent [19]

McMillan et al.

[11] Patent Number: 5,170,193

[45] Date of Patent: Dec. 8, 1992

[54] APPARATUS AND METHOD OF IDENTIFYING SIGNALS IN BIOLOGICAL TISSUES

[75] Inventors: **Charles F. McMillan**, Livermore; **H. Alfred Sklar**, San Francisco, both of Calif.

[73] Assignee: **Phoenix Laser Systems, Inc.**, San Francisco, Calif.

[21] Appl. No.: **656,722**

[22] Filed: **Feb. 19, 1991**

**Related U.S. Application Data**

[63] Continuation-in-part of Ser. No. 456,109, Dec. 22, 1989, Pat. No. 5,054,907.

[51] **Int. Cl.**[5] ..... **A61B 3/10**

[52] **U.S. Cl.** ..... **351/212**; 351/247

[58] **Field of Search** ..... 351/212, 247; 356/376

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,312,574 1/1982 Wilms ..... 351/212

4,878,750 11/1989 Sekiguchi ..... 351/212

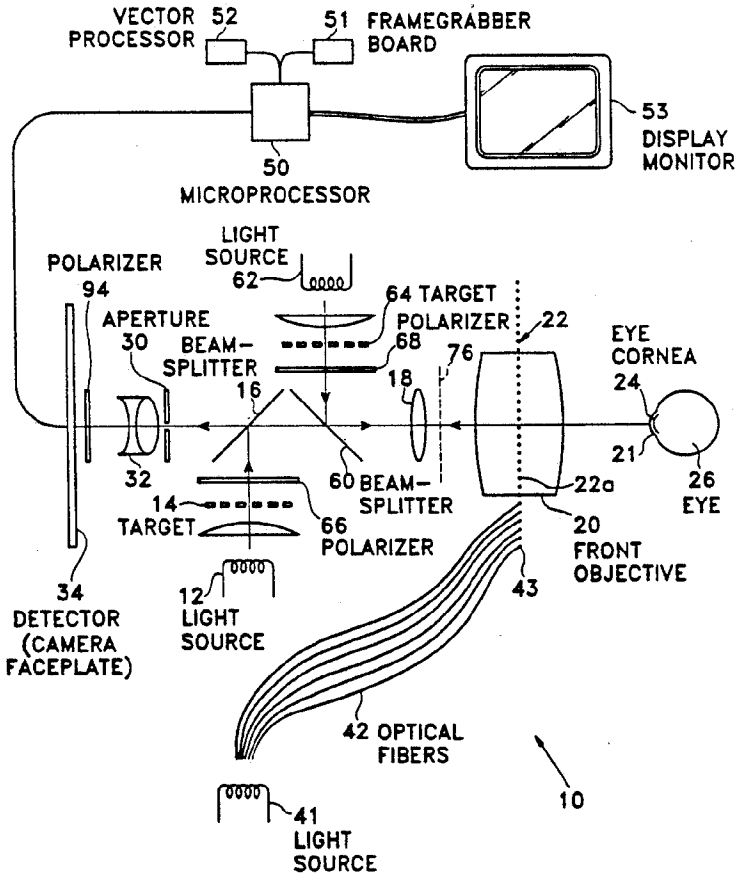
*Primary Examiner*—Rodney B. Bovernick

*Attorney, Agent, or Firm*—Thomas M. Freiburger

[57] **ABSTRACT**

The present invention describes a technique and apparatus for finding spots in an image with substantial noise making it difficult to identify without specialized noise suppression algorithms. In the context of determining corneal shape, as an example of the technique, the reflections of point light sources in or on the cornea have long played a diagnostic role. The image analysis technique described applies the tools of mathematical morphology and prior information about the shape of illumination patterns to remove noise and isolate the points of interest for further mathematical analysis. The output from the technique is a set of pairs matching the detected points in the image with the known location of the illumination.

**17 Claims, 7 Drawing Sheets**



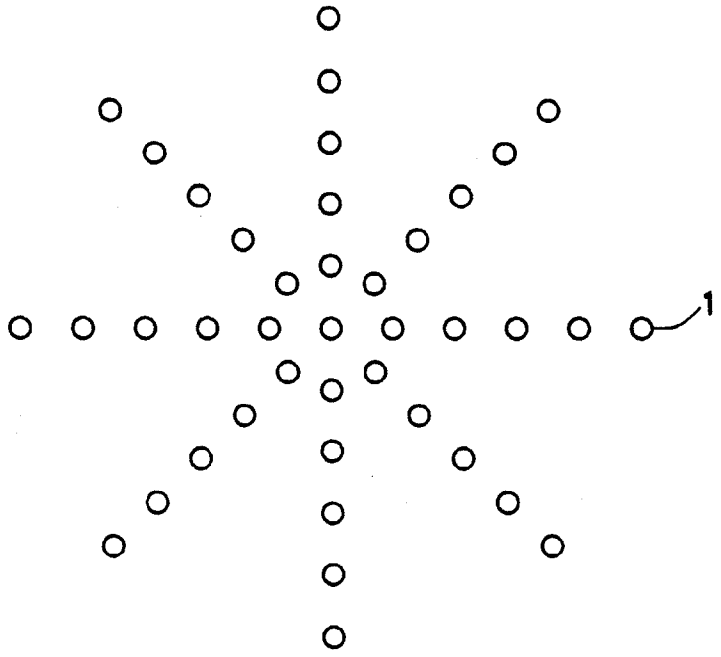



Fig. 1a

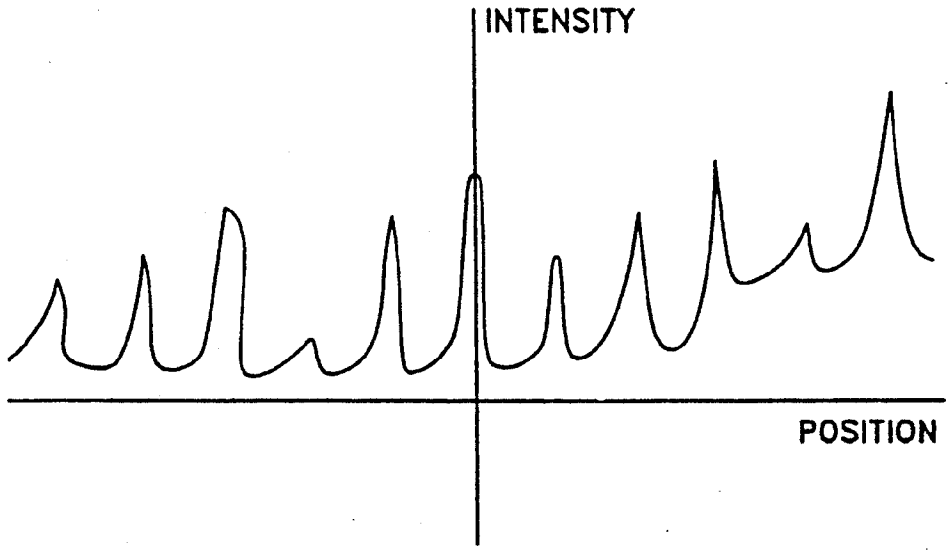



Fig. 1b

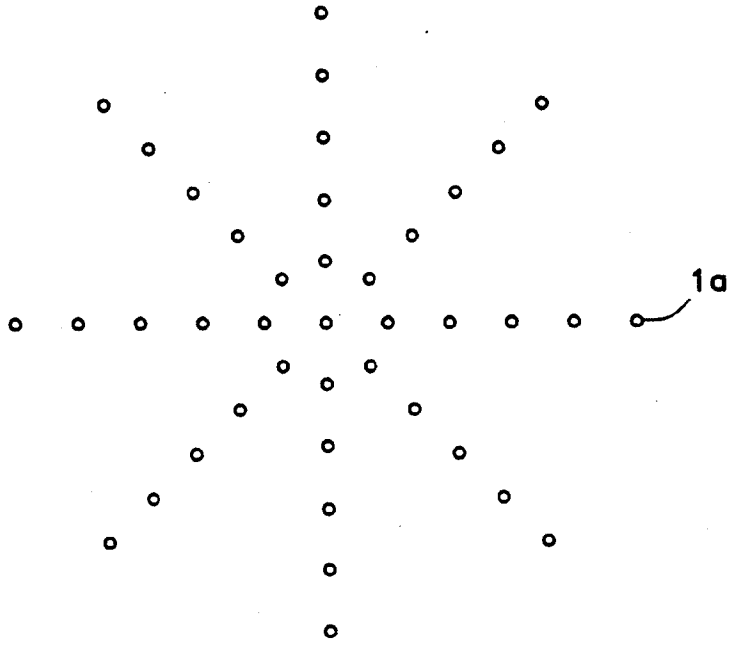



Fig. 2a

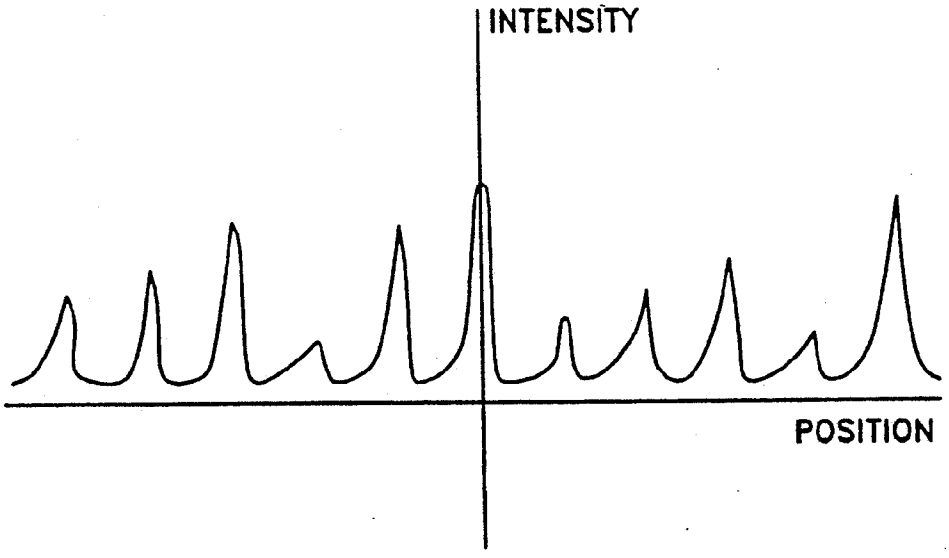



Fig. 2b

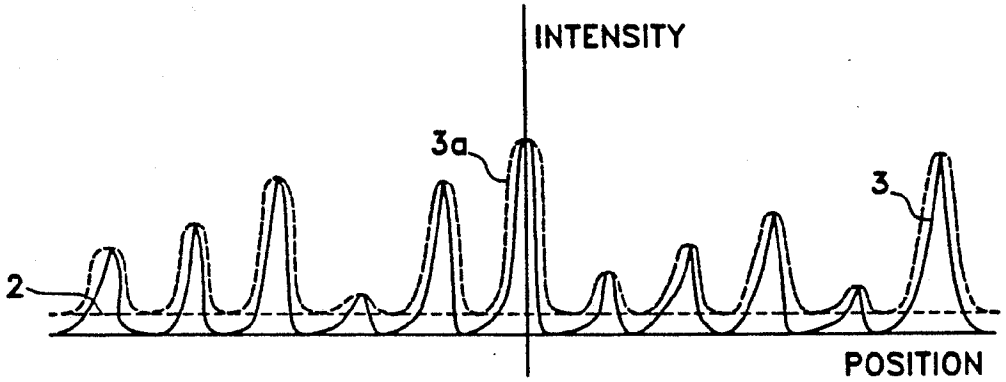



Fig. 3a

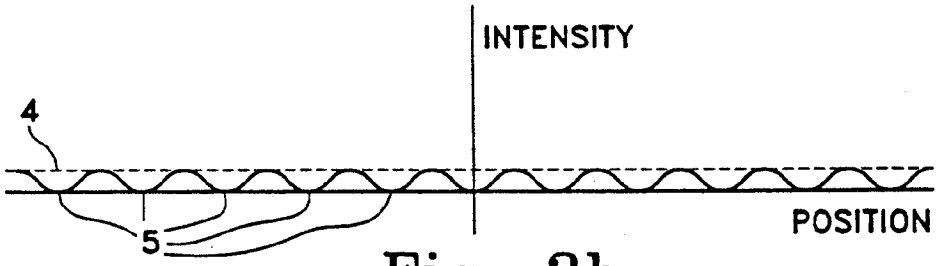



Fig. 3b

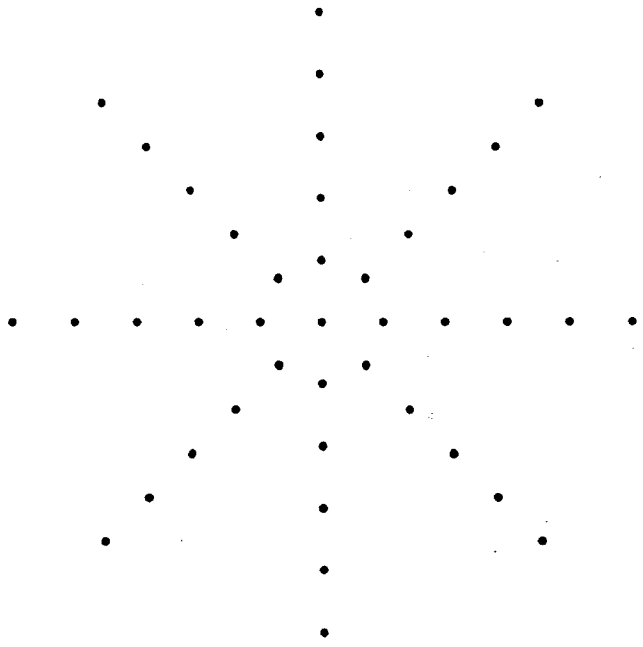


Fig. 4

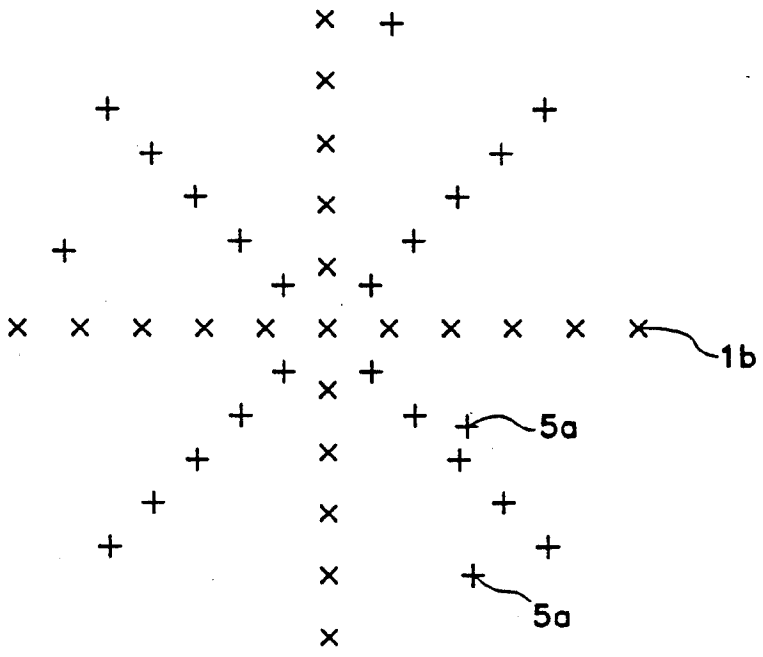



Fig. 5a

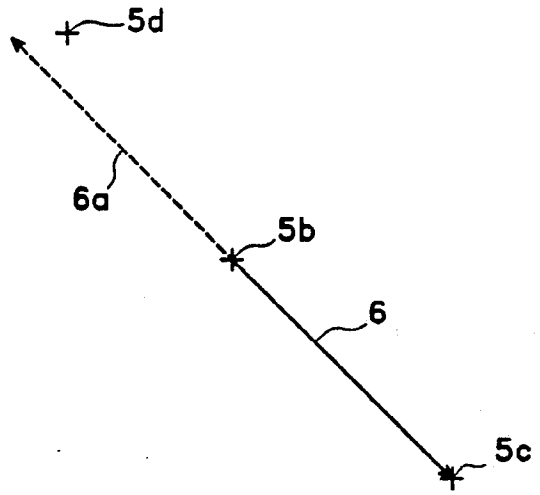



Fig. 5b

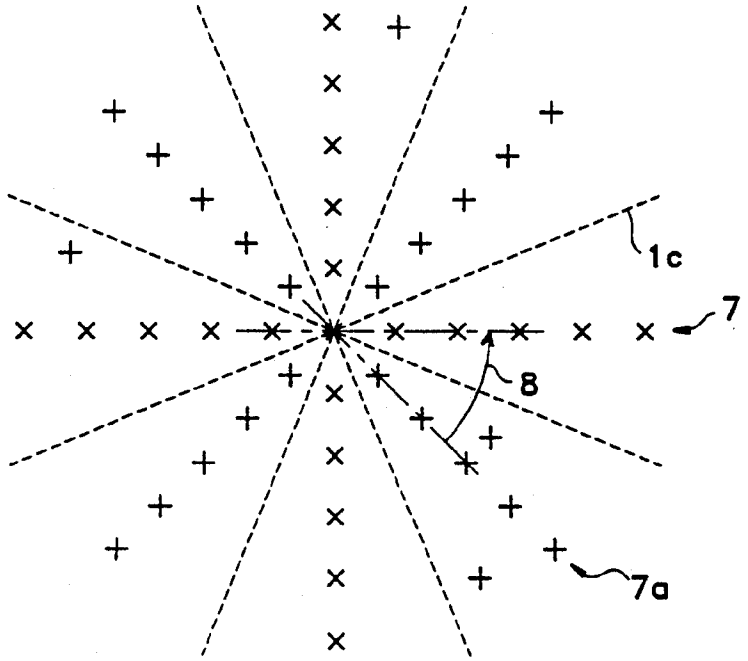



Fig. 6a

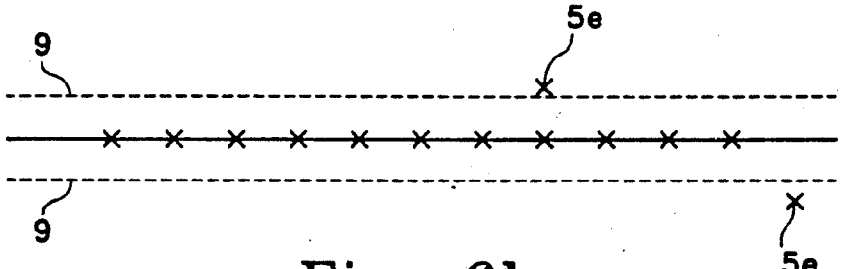



Fig. 6b

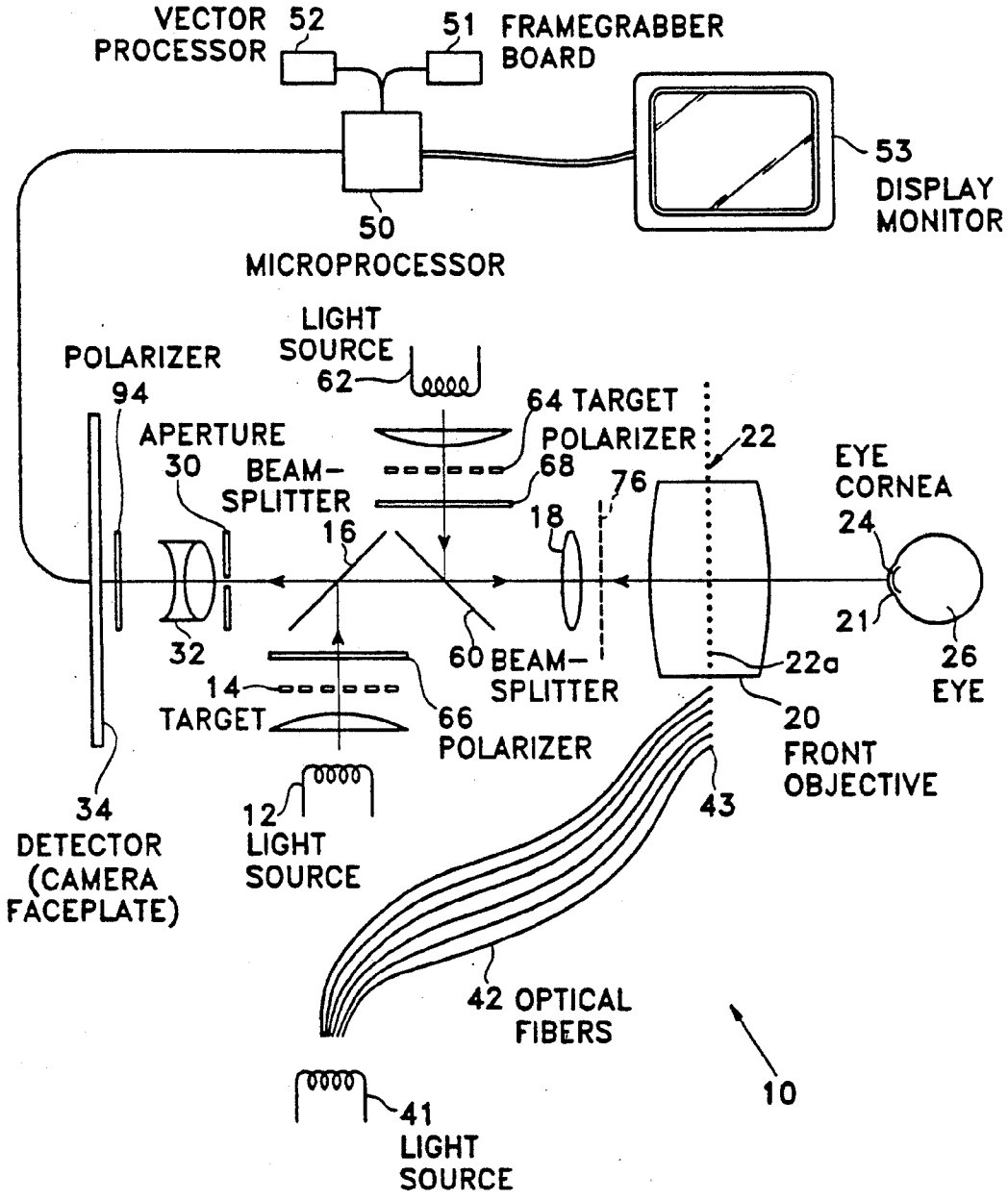

VECTOR PROCESSOR 52
51 FRAMEGRABBER BOARD
53 DISPLAY MONITOR
50 MICROPROCESSOR
LIGHT SOURCE 62
POLARIZER 94
APERTURE 30
BEAM-SPLITTER 16
64 TARGET POLARIZER 68
76
18
22
EYE CORNEA 24
21
26 EYE
22a
20 FRONT OBJECTIVE
43
32
14 TARGET
60 BEAM-SPLITTER
66 POLARIZER
34 DETECTOR (CAMERA FACEPLATE)
12 LIGHT SOURCE
42 OPTICAL FIBERS
41 LIGHT SOURCE
10

Fig. 7

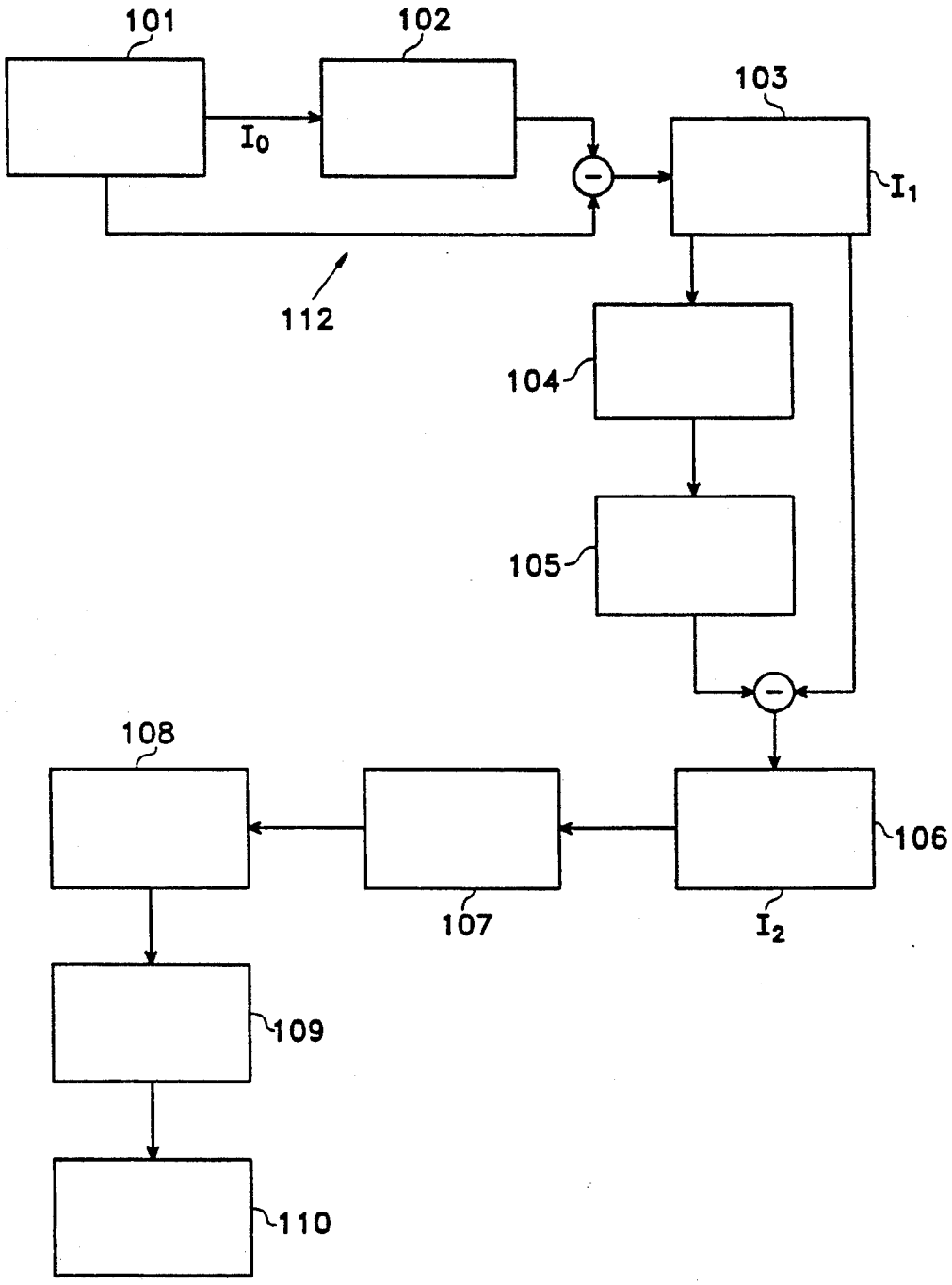



Fig. 8

# APPARATUS AND METHOD OF IDENTIFYING SIGNALS IN BIOLOGICAL TISSUES

## REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of copending U.S. application Ser. No. 456,109, filed Dec. 22, 1989, now U.S. Pat. No. 5,054,907.

## BACKGROUND OF THE INVENTION

The invention relates generally to optics, and more particularly to an imaging system wherein noise filtering techniques are employed to increase the signal to noise ratio used in analyzing the image.

In many biological systems, a method of preference for identifying the location, shape, or other characteristics of a tissue is to shine light or some other form of radiation onto the tissue and observe either how the light is reflected, absorbed, refracted, or scattered by the tissue. Inherent to observing the radiative effect on said tissue is the ability to detect a signal emanating from the tissue as a consequence of disturbing the tissue in some designated fashion. Many biological structures, however, are very sensitive to light or other forms of radiation thus limiting the threshold of signal that can be used to generate nondamaging responses. Even more troublesome is that many biological tissues have weak reflectivity or weak scattering properties so that the pertinent signal resulting from the disturbance is difficult to detect. This presents the observer with the compound problem of having to detect weak signals and low signal to noise ratios.

This is illustrated herein by discussion of the anterior surface of the eye's cornea known as the epithelium, although this discussion applies to any reflecting surface within the eye. When trying to define the surface topography of the epithelium, it is often desirable to shine point sources of light onto the epithelium from a precisely established location and to measure very accurately on a detector the location of the reflected image from said point source. However, point sources are a distribution of energy about some given location and said distribution is not generally constant or uniform from source to source. Since these energy distributions are reflected from unspecified surfaces at unpredetermined angles of incidence and with varying reflectivity and scattering, the identification of the reflected images of the point sources can be a difficult problem, especially when low illumination levels are desired.

The accuracy of the technique used by Sklar et al. to measure the surface profile of the cornea (U.S. patent application Ser. No. 456,109, now U.S. Pat. No. 5,054,907) hinges on the ability of the apparatus to measure the location of the rays from point sources in front of the eye as they are reflected from the cornea. As described by Sklar et al., the reflected rays are detected by an imaging device such as a CCD camera and digitized using a frame grabber card in a computer. Other perimeter devices used to measure eye features also rely upon accurately measuring the response of the eye to shining light onto the eye. In general, the observed reflections of these spots of light are embedded within a nonuniform background which may be very close to the noise floor level for contrast selectivity or threshold filtering because of poor light source edge definition and because of low reflectivity from the corneal epithelium.



The process of separating the spots of reflected signal from the background light and locating the peak or center of the signal presents a challenging image processing problem which can be crucial to the success of the entire surface profiling procedure. The apparatus and methods for obtaining the accurate location of such signals are the subject matter of this invention. These techniques are by no means restricted to ophthalmic applications, or even biological systems. These types of techniques have been in use in military applications where targets are at times difficult to differentiate from the background field and where filtering techniques such as thresholding and fast Fourier transform filtering prove unsuccessful. In medicine and, in particular, in surgery, computers have only recently begun to be incorporated as part of surgical devices. Since the techniques that are the subject of this invention require substantial calculation, these techniques have only recently become possible in a surgical environment.

As described further below, the apparatus and methods that are the subject of this invention use nonlinear filtering techniques of mathematical morphology many of which originated with the pioneering work of J. Von Neumann for developing automated devices to analyze images by comparing a given pixel of the image with its immediate neighbors. Many of these nonlinear filtering techniques are described by Serra (*Image Analysis and Mathematical Morphology,* Jean Serra, Academic Press 1982) and borrow from the fields of algebraic topology, harmonic analysis, stochastic processes, integral geometry, and others. We apply these nonlinear filtering techniques of mathematical morphology to the problem of isolating and identifying the centers of reflected point light sources. In the case of the invention described by Sklar et al. (U.S. patent application Ser. No. 456,109, now U.S. Pat. No. 5,054,907) the purpose for identifying said reflected light sources, herein after referred as "points", is to present the data to a surface profiling algorithm to determine the topography of the corneal surface. In other uses of the present invention not only different structures are to be described, but different types of image sources and scattering as well as absorptions, reflections, or refractions can be considered.

The problem of identifying the location of the centers of the points is addressed in three stages. First, the points must be isolated. Then we must look for and recognize symmetry in the pattern of points. And finally, we have rejection of noise pulses which have eluded previous elimination criteria.

## SUMMARY OF THE INVENTION

In accordance with the present invention, an apparatus and method for identifying the location of the centers of signals with poor signal to noise ratios and poor background differentiation are described. The method calls for first isolating the signal "points", recognizing the symmetries in the point patterns, and suppressing noise signals which have eluded previous elimination.

Since the recorded image background is frequently nonuniform due to scattering from the instrument, dust particles in the air, or even from the target, the first step in isolating the point signals is subtraction of the background. This is accomplished by subtracting a morphological gray-scale opening of the image using a circular kernel, which corresponds to an erosion followed by a dilation, from the original picture. The terms "erosion", "dilation", "opening", "closing", "morphological operations", "circular kernel", "hit or miss topology" and so

forth are intended in their algebraic sense as defined by J. Serra as cited above. For example, given two finite sequences $G_1, \ldots, G_i \ldots, G_m$ of open sets and $K_1, \ldots K_h, \ldots, K_n$ of compact sets in a real space of N dimensions, the class of all closed sets which hit every $G_i$ and miss every $K_h$ defines an open neighborhood in the space F of all closed sets in a real N dimensional space.

In the present invention, the size of the kernel is chosen so as to remove all features smaller than the approximate size of the dots from the image. The size of the dots is a known quantity since it is related to the illumination source. In the embodiment of the invention described below this corresponds to using a circular kernel of radius 3 pixels since the point sources are being magnified by the internal optics of the system to reflections which appear of that size to the detector observing the reflections. This first level of filtering of the original image, hereinafter referred to as $I_0$, shown in FIG. 1, leaves the resultant image, hereinafter referred to as $I_1$, shown in FIG. 2, with a level background and tends to equalize the maximum light intensities observed from the dots by reducing the more intense spots proportionately more than the fainter spots.

The next step in the process of isolating the points is to locate the peak intensities of the spots. This operation involves a combination of dilation with thresholding. A copy of the result of the first level of filtering described above, $I_1$, is stored in computer memory while a second image, $I_2$, is generated by applying a threshold transformation to $I_1$. As a consequence of this threshold filtering transformation, all pixels that composed $I_1$ which were below the threshold value are set to the threshold value while those above the threshold are left unchanged.

The threshold value is determined experimentally and is determined according to the types of light sources, optical elements, and targets to be identified. The threshold value is chosen to be low enough to be below the expected intensity level of the spots originating from the source light pattern, but above most of the observed random noise emanating from the floor. A dilation is then applied to the image $I_2$ once again using a circular kernel. The effect of this second filtering operation is to leave the maximum intensity of each spot unchanged and at the same location while spreading said maximum value of the spot around the peak location outwards to the full extent of the radius of the circular kernel. These peak values are determined by subtracting $I_1$ from $I_2$ and looking for areas in the resulting image whose intensities are identically zero. These areas of zero intensity correspond to the regions in image $I_2$ that were above threshold and represented local intensity maxima within the region determined by the radius kernel since they were unchanged between the original image $I_0$ and the dilated image. The effect of these operations is shown in FIGS. 3 and 4.

The choice of kernel radius for the dilation kernel must be based on the separation distance between adjacent spots in the light source which gave rise to $I_0$ and the anticipated smoothness of each of the intensity peaks observed. In practice, an illumination mask or a pattern of light sources is used with a given separation distance between the originating light spots. In ophthalmic applications where you are investigating the shape of the cornea, the distance between light spots is selected to correspond to the desired resolution as projected through the optical instrument of the apparatus onto the 12 millimeters which defines the diameter of the cornea taking into account the magnification of the apparatus as seen by the detector. In practice, the kernel radius should not exceed approximately half the average separation between intensity peaks. We say average because in general a uniform distribution of the peaks is neither required or optimally desirable.

For reasons of computational speed, it is preferable to make the kernel radius as small as possible. One of the limitations to reducing the kernel radius is that if the peaks are too jagged, the difference operation described above will reveal a variety of local minima rather than a single point at the peak of the spot whenever the smaller peaks are separated by more than the kernel radius. This proliferation of intensity peaks leads to confusion and degradation of the result. Fortunately, we can test for adequate choices for the kernel radius during experimental phases of setting up the apparatus for a given application in order to avoid the problem of proliferating peaks. Alternatively, the apparatus is set up to allow for modification of the kernel radius whenever the resulting image definition is insufficient because of the detection of excessive number of intensity peaks. For corneal topography applications we use a kernel of radius 5 pixels.

The resulting image $I_2$ described above as a result of the morphological transformations is to produce a list of points in $I_2$ made up primarily of the light rays originating in $I_0$ and reflected from the surface of the cornea. It is anticipated that $I_2$ will also contain some spurious noise spikes.

The next step in the process is to use knowledge of light source distribution symmetries to further suppress noise in $I_2$. The operation now is to find the center of an approximately symmetrical pattern so that it can be used to segregate a number of lines that pass through this center into separate groups. We say approximately symmetrical patterns for two principal reasons. First, the optics of the system and the reflecting surface itself will introduce some level of pattern distortion into $I_2$. Second, there is no requirement that we restrict discussion to only one pattern or even one type of pattern. Thus, to simplify the analysis a sorting into distinct pattern groups is desirable. An example of how these points might be distributed is shown in FIG. 5*a*.

In one embodiment of the present invention, the ophthalmic application described herein, a star shaped distribution of light sources being projected through the objective lens of a surgical workstation (Sklar et al, U.S. patent application Ser. Nos. 475,657 and 456,109, now U.S. Pat. No. 5,054,907) was selected as the desired mask in order to provide high definition of corneal curvatures in the neighborhood of the visual axis of the eye as well as a measure of angular asymmetry.

Even though the light source or input template used to produce the illumination pattern is symmetrical in this embodiment of the invention, the output pattern as observed by the detector is typically not symmetrical. Variations in illumination intensity which give rise to asymmetries can be caused by difference in fiber coupling the light source to the fiberoptic waveguides used to establish a pattern, or differential losses within the waveguide, or pinhole size variations is fabricating a mask for the illumination source. Such variations can lead to loss of detectable spots in the image $I_2$. Other mechanisms for spot loss can be ascribed to variations in reflectivity from the target being examined and even from image processing. The noise spots that have either

been created through the image processing or which have survived the various filtering operations described above will not generally be distributed symmetrically. Even more important, the distribution of noise spots are unlikely to correspond to the preselected source symmetries even when distorted. This information is used in further suppressing noise spots.

The process for finding the center of the pattern begins by going through the pattern and assuming that each peak in the pattern is the center of the pattern. The assumption that each peak is the center of the pattern is then tested by examining the reflection of each other spot in the pattern about the supposed center and searching for existence of another spot close to the reflection. Namely, the vector from the assumed center to the test point is reversed. If there exists another point near the reflection, the given score scale value for that particular assumed center is increased. After sequentially going through this process of testing each peak for closest neighbors with every surviving spot in $I_2$, the peak with the highest score corresponds to the center. This process is illustrated in FIG. 5*b*.

This process of locating the center is calculationally time consuming. It is, however, very robust. It successfully locates the center of a given illumination pattern as reflected by the eye even when many spots are missing or with a multiplicity of extraneous noise spots. When dealing with the variety of ailments which are often encountered by ophthalmologists, keratoconus being one notable case, surface aberrations can frequently cause illumination spots to go astray. A robust system of detection which will allow an instrument to perform irrespective of such variation was one of the motivations in the design of the present invention. With the computational capacity available today with high performance dedicated computer boards using 80386 and 80486 based microprocessor with specialized mathematical coprocessor chips and vectored accelerator boards, the calculational time of the processes described herein which were formerly practicable on mainframe computers become accessible to commercial uses and surgical environments.

The template pattern **46** described in FIG. **9** and the pattern **82** in FIG. **11** of Sklar et al. (U.S. application Ser. No. 456,109) correspond to a sequence of light sources arranged along a number of intersecting straight lines. If we apply the above general discussion to these templates, the effort in finding the center of the pattern reduces to locating the intersection point for the straight line patterns as observed in image $I_2$. Once the center of the pattern is established, the approximately straight lines of spots radiating outward from this center in image $I_2$ are easily identified by collecting spots together that have approximately the same angle relative to the axis of the line based upon the arctangent value at the given spot.

The grouping of spots into independent straight lines can be based upon the a priori knowledge of the number of such straight lines which composed the initial illumination pattern as described in FIG. **9** and **11** noted above. Thus, the correct number of angular bins can be chosen for the collection system as well as reasonable values for the boundaries of the bins. In other words, we are using knowledge of the shape of the illumination pattern in order to improve noise suppression in a system with poor signal to noise ratios.

In our illustrative example, we use eight straight lines of illumination source points passing through the center shown in FIG. **6**. We thus divide the spots observed in $I_2$ into eight separate bins using angles midway between each known line as the bin boundaries. So far, we have used the knowledge that the illumination light sources have been arranged along discrete and distinct patterns and we have used the existence of such patterns to locate the center of the pattern and to sort the observed spots into distinct bins or subpatterns. We now proceed to use knowledge of the original shape of the distinct illumination pattern to further suppress noise.

The final noise rejection operation in the system is based upon a priori knowledge of the configuration of the illumination system. In the example noted above concerning topographical mappings of the eye's cornea, the subpatterns described in FIGS. **9** and **11** correspond to straight lines intersecting at some center point. The angles of each line with respect to one another are predetermined during fabrication of the illumination masks. Using this information, the spots in image $I_2$ were segregated into individual groups or bins. Each grouping of points in a bin can then be approximately rotated to the horizontal axis by making use of the knowledge of the angular relation between each line in the illumination mask.

Any point within the group lying more than a preselected number of pixels off the horizontal axis following the rotation is identified as a noise pixel and deleted from the image. This operation is shown in FIG. **6**. In practice, we have found that a discrimination value of 5 pixels is adequate for the task of noise suppression. After elimination of such noise pixels, the subpatterns can be counterrotated to its corresponding alignment to form image $I_3$. In general, there is no requirement that the subpatterns correspond to straight lines or that the superposition transformation be isometric. It is important however that the patterns be selected so as to permit superposition transformations which are invertible.

Once the points that are well away from the true line have been eliminated, the resulting pattern $I_3$ is compared with a stored calibration pattern where the scales of the two patterns are matched by comparing the average spacing between points in each line subpattern. Points which do not have a mate in the calibration pattern are further ignored before proceeding to the profiling procedure. Examples of such profiling procedures are discussed in Sklar et al U.S. Pat. No. 5,054,907.

## DESCRIPTION OF THE DRAWINGS

FIG. **1***a* shows the illumination pattern $I_0$ of the light reflected from the cornea or from the tear layer overlying the cornea as digitized by the frame grabber. A trace across the pattern at **1** shown in FIG. **1***b* illustrates the nonuniform peak heights and background.

FIG. **2***a* shows the illumination pattern $I_1$ after the background has been subtracted. FIG. **2***b* shows a trace across the pattern illustrating the equalization of the background level.

FIG. **3***a* shows a trace **3** of image $I_1$, a reasonable threshold value 2, and a trace **3***a* across the dilated threshold image. The difference between trace **3** and trace **3***a* is shown in FIG. **3***b* with trace **4** representing the threshold level and the points **5** representing the points at which the difference is identically zero indicating the local maxima of trace **3**.

FIG. **4** shows the resulting image after the difference operation with the threshold set so that only the null points are visible.

FIG. 5*a* shows the distribution of points found in FIG. 4. The points labeled 2 represent noise signals. FIG. 5*b* illustrates the process of finding the point about which the pattern is symmetrical. A trial center point is denoted by 5*b*. The vector 5 denotes the vector from the trial center point 5*b* to a given test point 5*c*. The vector 6*a* depicts the reflection of vector 6 about the trial center 5*b* and the point 5*d* is a point near the reflected vector 6*a*, thus incrementing the score of point 5*b*.

FIG. 6*a* shows the pattern of dots with angular bins as indicated in trace 1*c*. The horizontal axis is 7. To remove noise spots from one of the lines, say line 7*a*, the pattern 1*c* is rotated through angle 8 to give the result shown in FIG. 6*b*. The lines 9 show the discrimination threshold, in our case 5 pixels, and the points 5*e* represent points that are rejected since they lie sufficiently far off line.

FIG. 7 shows an embodiment of an apparatus which illuminates the target and provides the optical means for detecting the signal and sending it the microprocessor means that initiates the noise suppression techniques discussed herein.

FIG. 8 is a schematic drawing of the information flow for the filtering techniques that are a part of the present invention. The frame grabber board **101** inputs the image $I_0$ to the image opening operation **102**. We then develop at **103** the image of $I_0$ with the background removed. The thresholding operation **104** replaces all intensity values less than the threshold by the threshold value and is followed by the dilation operation **105**. The pattern **106** selects the points that are identically zero representing the local maxima in $I_0$. The process of finding the center of the pattern **107** uses symmetry scoring. Next follows the process **108** of segregating the spots into lines based on the a priori information about the angles of the lines in the illumination source. The process **109** of rejecting the spots which are off the line after rotation then follows. The final step **10** denotes the process of matching the spots in the image lines to spots on the target lines. Non-matching spots are deleted.

## DESCRIPTION OF PREFERRED EMBODIMENTS

In the drawings, FIG. 7 shows in schematic representation an example of how an instrument would make use of the noise suppression features and methods of the present invention. It describes a system of optical elements in accordance with the invention for use in carrying out ophthalmic diagnosis and analysis similar to a device taught by Sklar et al. (U.S. patent application Ser. No. 456,109).

The system, generally identified by the reference number **110**, includes an illuminator or light source **12**, a pattern plate or disk **14** having a pattern of holes cut in the plate for producing a desired pattern of discrete light sources, a non-distorting beam splitter **16**, a lens **18** which projects an image of target **14** onto an image plane at **22**. This image plane **22** is close to or coincidental with the system of objective lens **20**. The purpose of placing the image at this location **22** is to have the objective lens **20** serve as a field lens, that is bending the rays of light **22***a* that form the image towards the patient's cornea **21**.

As indicated in FIG. 7, the focused image **22** of the pattern is a real image, formed at some plane at or near the lens **20** and between the lens and the patient. The real image preferably is in the lens **20**, but it can be very closely in front of the lens (i.e. a few millimeters in front) or behind the lens. The importance is that the focused image **22** lie along the optical path passing through the lens **20** which can give optimal definition along the visual axis of the cornea. In this real image, each point source of light **22***a* projects a cone of light towards the patient. Thus, each point source **22***a* in the real image makes an infinite number of specular reflections off the front surface of the cornea **24** of a patient's eye **26**. In addition to the focused image **22**, it may prove desirable to extend the range of coverage of a curved reflecting surface **21** by projecting a longer array of image points **22***a* than can be projected through the final lens **20**.

Towards this end, a light source **41** illuminates a fiberoptic bundle **42** whose polished terminations **43** serve as additional image points to the image points **22***a* located within or directly in front of the final lens **20**. The purpose of light points **43** is to increase the length of the image line **22** and thus provide more extensive peripheral definition of cornea **21** away from the visual axis of the eye. The position of light points **43** is adjusted so as to lie along the plane of pattern **22**.

Alternatively, even for conventional perimeter devices that only utilize light sources as described by light points **43**, the present invention provides a means for improving the detection of said light points **43** when poor signal to noise ratios are perceived by the detector **34**. This enables weak light sources **12** and **62** to be utilized without sacrificing resolution, and thus minimize discomfort to the patient or light toxicity to the patients eye.

The F-number of the final lens **20** determines the maximum area of the cornea that can be measured. The objective lens serves as a field lens, and the patient's cornea must be at the focal length of the lens **20**. A preferred method for assuring that the eye is located at the focal length of the lens **20** is discussed in Fountain et al. (U.S. patent application Ser. No. 655,919). This assures that the light reflected off the eye parallel to the optical axis of the instrument is then brought to a point behind the lens **20** at the focal distance of the lens **20**. This enables the return light to be apertured down, to select only those rays which were paraxial off the eye. This enables the system to localize a detected point to a point on the cornea from which that ray is reflected. If the objective lens **20** were not situated to serve as a field lens, outermost points of light in the pattern would not reflect off the cornea. As a field lens, the lens **20** efficiently bends the outer points of light toward the eye.

It is preferred that the focal length of the lens be great enough to provide an unobstructed, comfortable distance from the instrument to the patient and adequate working room for the surgeon, for surgical applications.

FIG. 7 schematically indicate-s that the detector or camera plane **34** is connected to a microprocessor **50** which may contain a frame grabber board **51** which digitizes the images detected by detector **34** together with a vector processor **52** for speeding the calculations. The microprocessor may be connected to a display device, such as a CRT monitor **53** as indicated. Data gathered from the system as described is received by the microprocessor **50** and analyzed. As described below, each detected point is correlated with the location of the particular point in the source pattern from which it emanated. This information is then provided to an algorithm as described by Sklar et al. (U.S. patent application Ser. No. 456,109, now U.S. Pat. No.

5,054,907) for display as part of a user interface, for ophthalmic surgical procedures such as described by Sklar et al. (U.S. patent application Ser. No. 307,315 now U.S. Pat. No. 5,098,241 and Ser. No. 475,657) or for diagnostic verification of the eye's topography.

FIG. **8** shows a chart of the flow of information and the multiple steps taken in filtering the noise from an image such as those contemplated in the present invention. The chart, discussed above, traces the various mathematical morphological transformations which are applied to the images generated in the particular case of the configuration shown in FIG. **7** which constitutes an application in the field of ophthalmology.

As explained above, FIG. **8** describes the flow of information from the frame grabber **101**, noted as the frame grabber means **51** in FIG. **7**, to the profiling system. The image from the frame grabber **101** is shown in FIG. **1***a*. It is formed by the light from the spots **22** in FIG. **7** reflecting from the cornea or the cornea's tear layer, passing through the aperture, and hitting the detector. A trace across this pattern along the line shown as **1** in FIG. **1***a* shows the intensity along the line as a function of position. The result is shown in FIG. **1***b*. The pattern in FIG. **1***b* is referred to as the image $I_0$.

A copy of $I_0$ is passed through the morphological opening operation using a circular kernel shown as **112** in FIG. **8**. The difference between the original image Iand this opening is represented by **103** in FIG. **8**. We refer to this image as $I_1$. The effect of the opening and the difference is to equalize the background level throughout the image, thus effectively removing the background from the resulting image shown in FIG. **2***a* with a trace across the pattern illustrating the equalization of the background level when compared with FIG. **1***b*.

While a copy of $I_1$ is maintained, the information is passed first through a thresholding operation in which all pixel values in the image less than the threshold value are set to the threshold value, represented schematically as **104** in FIG. **8**, the output of which passes through a dilation operation **105** in FIG. **8**. The effect of these operations is shown in FIG. **3***a*. In FIG. **3***a*, the line **3** shows a trace across $I_1$, and **2** represents a reasonable threshold level. Trace **3***a* shows what the trace looks like after both thresholding and dilating. The difference between $I_1$ and the thresholded and dilated image is noted in **106** of FIG. **8**. This difference is referred to as image $I_2$ and is shown in FIG. **4**. The places in $I_2$ that are identically zero represent the local maxima of $I_0$, which are the points sought by this technique. A trace across this pattern is shown in FIG. **3***b* with the points labeled **5** being the places that are identically zero in the pattern corresponding to the local maxima of **3** in FIG. **3***a*.

The collection of points in $I_2$ that are zero must then be organized for subsequent analysis. These points are shown in FIG. **5***a*. Because of noise in the original input image, there are still a few noise spots in the pattern illustrated by **5***a* in this figure. These points are eliminated by the following procedure. The first step in the process is to find the center of the pattern represented by **107** in FIG. **8**. The process assumes that each of the spots $I_2$ is the center of the pattern in turn and then examines the rest of the spots in the pattern to measure if they are symmetrical about the assumed center as illustrated in FIG. **5***b*. Point **5***b* denotes the assumed center. Point **5***c* is a test point. Vector **6** is the vector joining point **5***c* to point **5***c*. The operation of checking for symmetry involves reflecting the vector **6** about the point **5***b* and looking to see whether there is another point in the pattern near this reflection of the vector. In the case depicted, there is indeed a point nearby, namely point **5***d*, so the symmetry score for point **5***b* would be incremented. After using all the remaining points in the pattern as test points for the assumed center **5***b*, we would arrive at a total symmetry score for the assumed center **5***b*. After checking each point in the pattern as an assumed center, we arrive at a symmetry score for each point. The point with the highest symmetry score is the actual center of the pattern. While the process of finding the center of the pattern is a very simple operation for the human eye, even when points are missing and extraneous points find their way into the field of view, this is a nontrivial process for a computer. The process described as part of the present invention is robust even in the presence of noise.

Once the center of the pattern has been identified, the angles of the pattern lines relative to the axis, known from the construction of the target, noted as **14** in FIG. **7**, can be used to segregate the dots into distinct lines. This process is depicted in **108** of FIG. **8**. FIG. **6***a* shows the pattern with segregated lines **1***c* drawn at the bisectors of the pattern lines. Any points that lie between adjacent segregation lines are considered to belong to the same line. The set of points **7** form the line that is on the horizontal axis of the system, while a line **7***a* corresponds to a line off the horizontal axis at an angle $\Omega$, shown as **8**. Note that $\Omega$ is known from the fabrication of the target illuminator.

Each group of dots grouped as a line by process **108** of FIG. **8** is examined for extraneous dots. This is done in **109** of FIG. **8** by rotating each group through its corresponding angle $\Omega$ and deleting points that are more the discrimination value away from the horizontal axis. The discrimination levels **5** are used to keep points within the lines and delete points **6** which fall outside the criteria. This process is performed for each line in turn deleting points outside a given distance from the axis.

The final step in the process involves matching the detected points in the image with the known location of points in the illumination target **14** of FIG. **7**. This process is represented by **110** in FIG. **8**. Since the center point is known from process **107** of FIG. **8**, it is the first point to be matched. Since the known angle of each of the target lines was used to get rid of noise pulses during the rotation of image lines, it can also be used to tell which group of dots goes with which line of the illumination target **14**. While the magnification between the image and the target is unknown since it depends on the radius of curvature of the cornea, the average spacing of the dots on the illumination target line and the dots on the image line can be used to scale each line so that the two patterns are normalized. Thus the distances of the points on both the target and the image from the center are divided by their respective averages. The normalized points from both patterns are then examined to make sure that there is a corresponding point within a sufficiently small range chosen to be small compared to the normalized dot spacing. Any point in either the image or the target pattern that is found to be without a corresponding mate is discarded before the matched data set is sent to the profiling algorithm described by Sklar et al. (U.S. Pat. No. 5,054,907).

The above described preferred embodiments are intended to illustrate the principles of the invention, but

not to limit its scope. Other embodiments and variations to these preferred embodiments will be apparent to those skilled in the art and may be made without departing from the spirit and scope of the invention as defined in the following claims.

We claim:

1. An ophthalmic diagnostic instrument for determining the shape of the cornea, comprising,

an objective lens as an optical element of the instrument, on or symmetrically about an optical axis of the instrument,

means for projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located between the interior of the objective lens and the eye,

means for expanding the region of coverage on the cornea by using the objective lens as a field lens for the pattern image,

means for selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and for detecting a reflected position of substantially each point light source, as reflected from the cornea, including means for relaying the Fourier plane of the objective lens to a relayed position in the instrument, with aperture means positioned at said relayed position for limiting the collected light to that which is reflected paraxially off the cornea, whereby the aperture means is a spaced distance from the objective lens,

means for analyzing the returned, collected pattern image and for filtering the noise from the pattern image using mathematical morphological transformations,

means for comparing the filtered, collected pattern image to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected, and

means for deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image.

2. Apparatus according to claim 1, wherein the pattern of discrete separated point light sources includes more than one rectilinear sequence.

3. Apparatus according to claim 1, wherein the pattern of discrete separated point light sources includes an asymmetrical shape having a plurality of lines of point light sources.

4. Apparatus according to claim 1, wherein the pattern of discrete separated point light sources lie along an invertible function of distance from the optical axis of the objective lens.

5. Apparatus according to claim 1, in combination with a surgical microscope having a standard auxiliary camera mount, and the ophthalmic diagnostic instrument being connected to the surgical microscope via the auxiliary camera mount, with an objective lens of the surgical microscope serving as the objective lens of the ophthalmic diagnostic instrument.

6. Apparatus according to claim 1, wherein the ophthalmic diagnostic instrument includes means for folding the pattern of discrete separated point light sources onto the optical axis of the instrument, toward the cornea, with the means for projecting the pattern including a source of the pattern off-axis from the optical axis and from the path of the returned, distorted pattern image.

7. Apparatus according to claim 1, further including means for projecting a variety of predetermined light patterns whose reflected image can be individually mapped isometrically onto a straight line simultaneously with said pattern of discrete separated point light sources, and means for separately analyzing distorted reflected light from the cornea relating to the invertible shapes and for providing separate, qualitative information which can be compared with the corneal surface shape derived via the pattern of discrete separated point light sources.

8. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the back or endothelial surface of the cornea.

9. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the anterior surface of the eye's lens.

10. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the posterior surface of the eye's lens.

11. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the retina.

12. Apparatus according to claim 1, further including means for separately analyzing a secondary returned, reflected pattern image as reflected from the floaters in the eye's vitreous.

13. Apparatus according to claim 8, wherein said means for separately analyzing includes filtering means for electronically separating returned light points on the detector means occurring from the front surface of the cornea from those occurring from the back surface of the cornea, by separating different ranges of amplitude of the detected light.

14. Apparatus according to claim 1, wherein the means for projecting a pattern includes an illuminating light source and a plate with a laser-cut or photolithographically produced pattern of discrete holes to form the discrete separated point light sources.

15. A method for determining the shape of the cornea of an eye, comprising,

projecting a pattern of discrete separated point light sources and forming a real image of the pattern of point light sources at a position located in front of the eye,

selecting and collecting a reflected image of the pattern as reflected paraxially off the cornea, and detecting a reflected position of substantially each point light source, as reflected from the cornea,

analyzing the returned, collected pattern image, including filtering noise from the collected pattern image using mathematical morphological transformations, thresholding, or fast Fourier transformation techniques,

comparing the filtered collected pattern image to the undistorted pattern as projected, including analyzing the relative location and spatial orientation of the reflected point light sources as compared to the pattern as projected, and

deriving mathematically a close approximation of a corneal surface shape that would give rise to such a collected pattern image.

16. The method of claim 15, wherein the pattern of discrete separated point light sources comprises a generally asterisk shaped pattern with an intersection point

at the optical axis of the instrument, and including means associated with the pattern for establishing a readily identifiable rotational orientation of the pattern.

17. The method of claim 15, wherein the pattern of discrete separated point light sources can be represented as any uniquely invertible function of position with respect to the optical axis of the image detection means.

* * * * *

# APPENDIX D

US006050687A

# United States Patent [19]

Bille et al.

[11] Patent Number: 6,050,687

[45] Date of Patent: Apr. 18, 2000

[54] **METHOD AND APPARATUS FOR MEASUREMENT OF THE REFRACTIVE PROPERTIES OF THE HUMAN EYE**

[75] Inventors: **Josef Bille**; **Frieder Loesel**, both of Heidelberg, Germany

[73] Assignee: **20/10 Perfect Vision Optische Geraete GmbH**, Heidelberg, Germany

[21] Appl. No.: **09/332,297**

[22] Filed: **Jun. 11, 1999**

[51] **Int. Cl.**[7] ..... **A61B 3/10**

[52] **U.S. Cl.** ..... **351/212**

[58] **Field of Search** ..... 351/205, 206, 351/211, 212, 214, 221; 623/4, 5

[56] **References Cited**

U.S. PATENT DOCUMENTS

| | | | |
|---|---|---|---|
| 4,579,430 | 4/1986 | Bille . | |
| 4,732,473 | 3/1988 | Bille et al. . | |
| 4,772,115 | 9/1988 | Gersten et al. . | |
| 4,848,340 | 7/1989 | Bille et al. . | |
| 4,887,592 | 12/1989 | Loertscher . | |
| 4,988,348 | 1/1991 | Bille . | |
| 5,062,702 | 11/1991 | Bille . | |
| 5,156,622 | 10/1992 | Thompson | 623/5 |
| 5,512,965 | 4/1996 | Snook | 351/205 |

OTHER PUBLICATIONS

Liang, Junzhong et al, *Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave–front sensor, J. Opt. Soc. Am A,* vol. 11, No. 7, Jul. 1994, pp. 1949–1957.

Walsh, G. et al. *Objective technique for the determination of monochromatic aberrations of the human eye, J. Opt. Soc. Am A* vol. 1, No. 9, Sep., 1984.

Wang, J. Y., et al., *Wave–front interpretation with Zernike polynomials, Applied Optics,* vol. 19, No. 9, May 1, 1980, pp. 1510–1518.

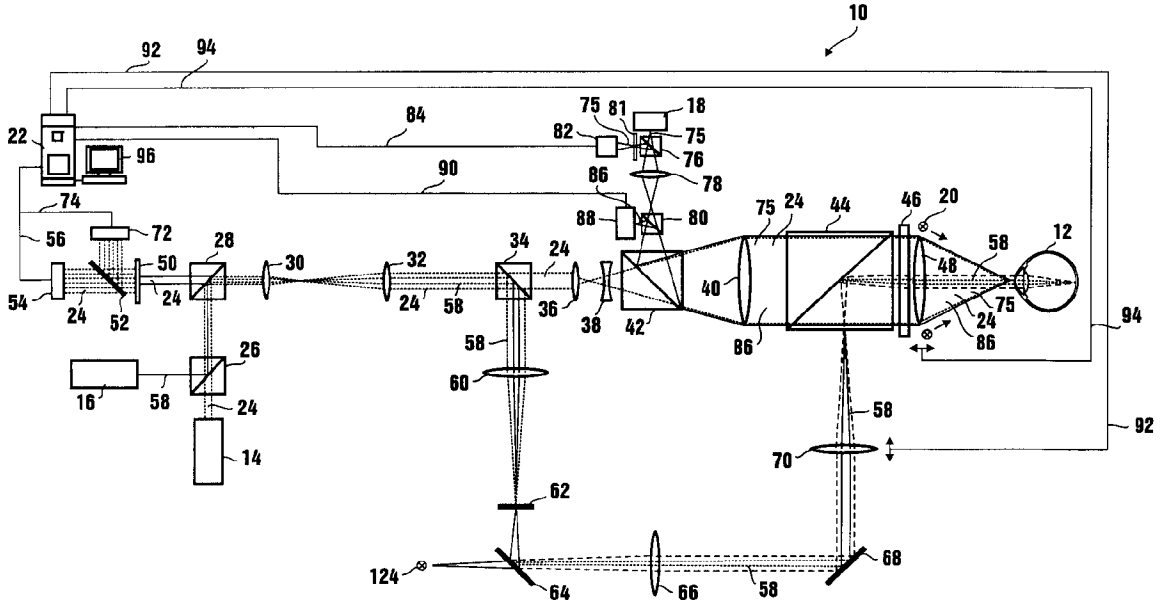
*Primary Examiner*—George Manuel

*Attorney, Agent, or Firm*—Nydegger & Associates

[57] **ABSTRACT**

A system, and its method, for measuring the optical properties of an eye employs optical components for detecting both the reflection of a first light beam from the anterior surface of the eye, and the reflection of a second light beam from the retina of the eye. Sensors are included to receive and separate each of these reflected light beams into a plurality of individual beams, each having its own specific optical path length. The optical path lengths of individual beams reflected from the cornea are collectively used to create a topographical map of the cornea's anterior surface and the optical path lengths of individual beams reflected from the retina are collectively used to create an acuity map of the entire eye. Further, there are additional optical components which respectively determine the position of the eye, a length for the eye and an aberration for the relaxed lens of the eye. A computer is then used to compare the topographical map with the acuity map, while compensating for the lens aberration, to construct a topography for the posterior surface of the cornea.

**20 Claims, 3 Drawing Sheets**



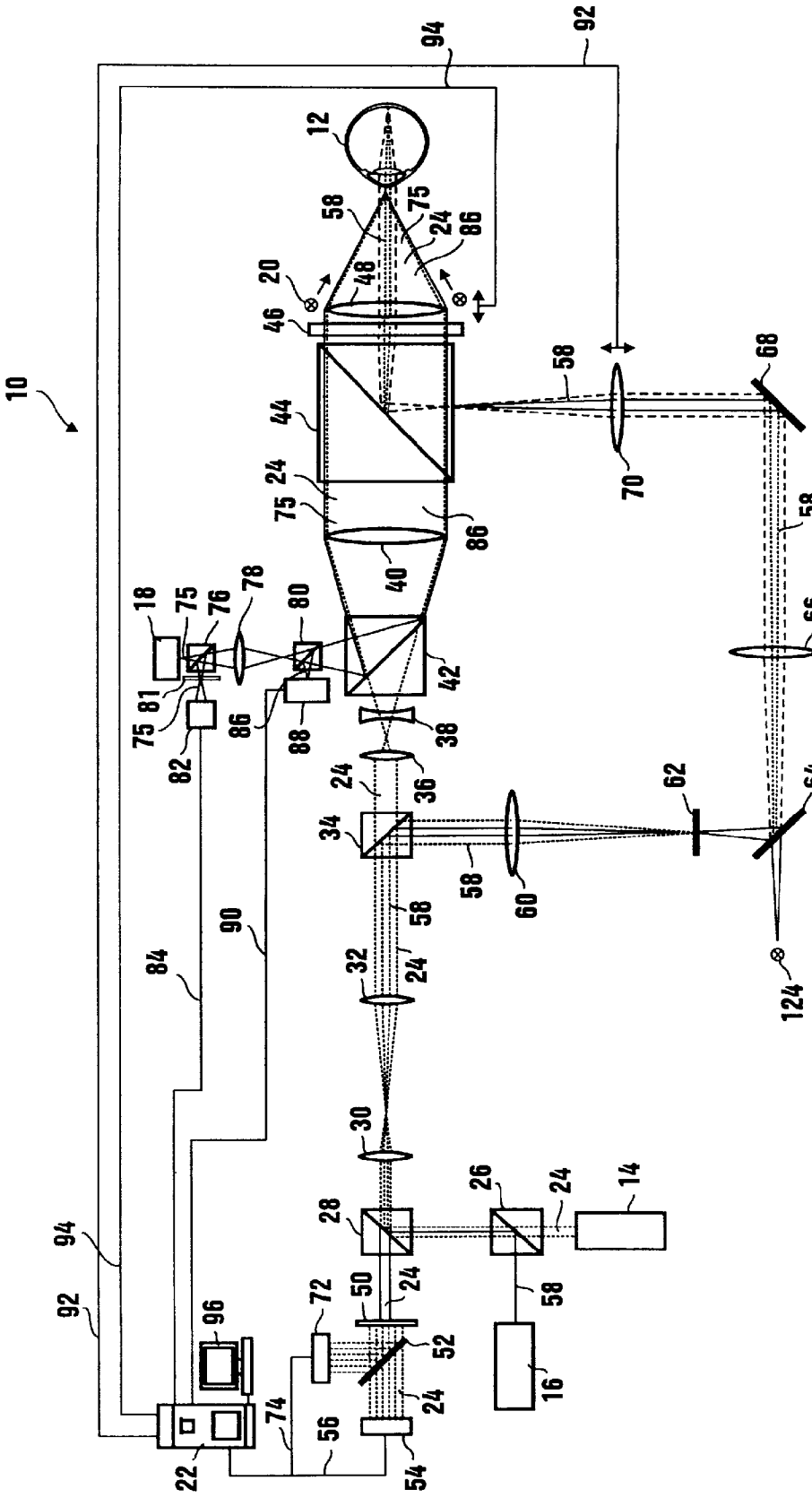

10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64
66
68
70
72
74
75
76
78
80
81
82
84
86
88
90
92
94
96
124

Figure 1

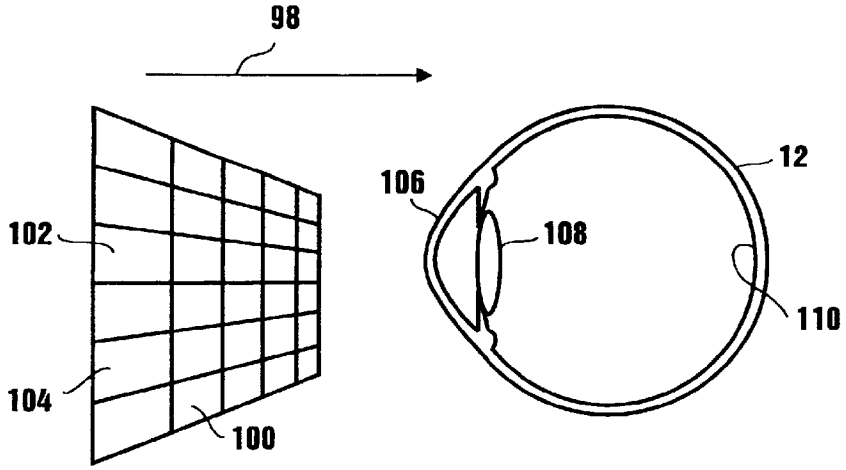



Figure 2A

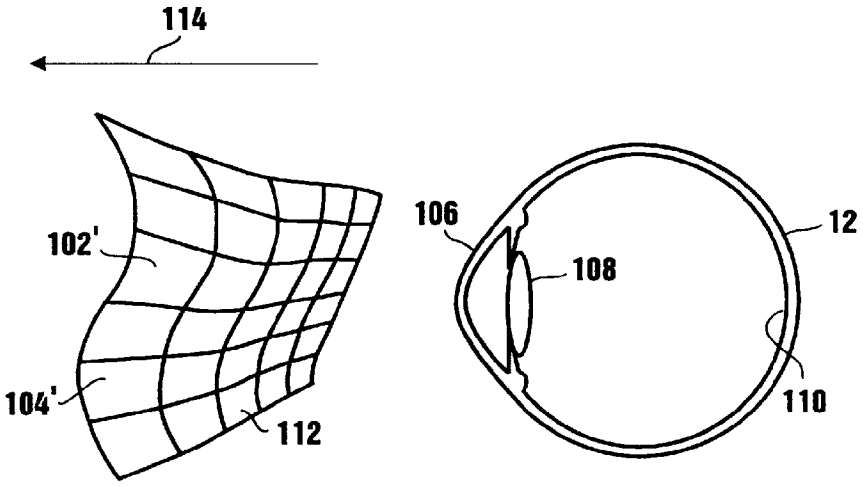



Figure 2B

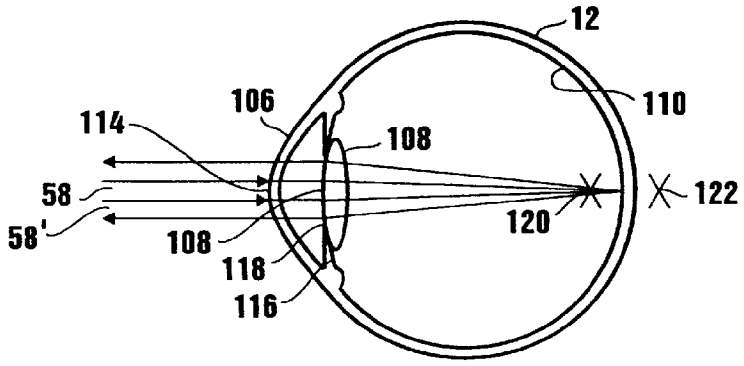



Figure 3

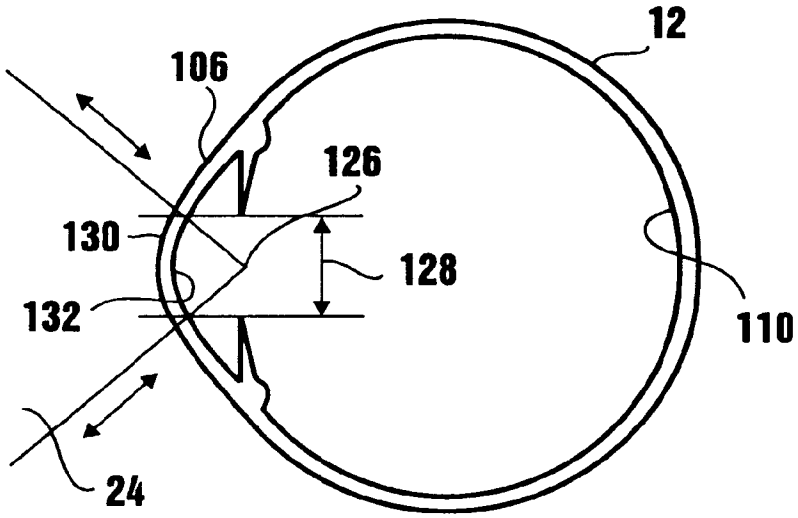



Figure 4

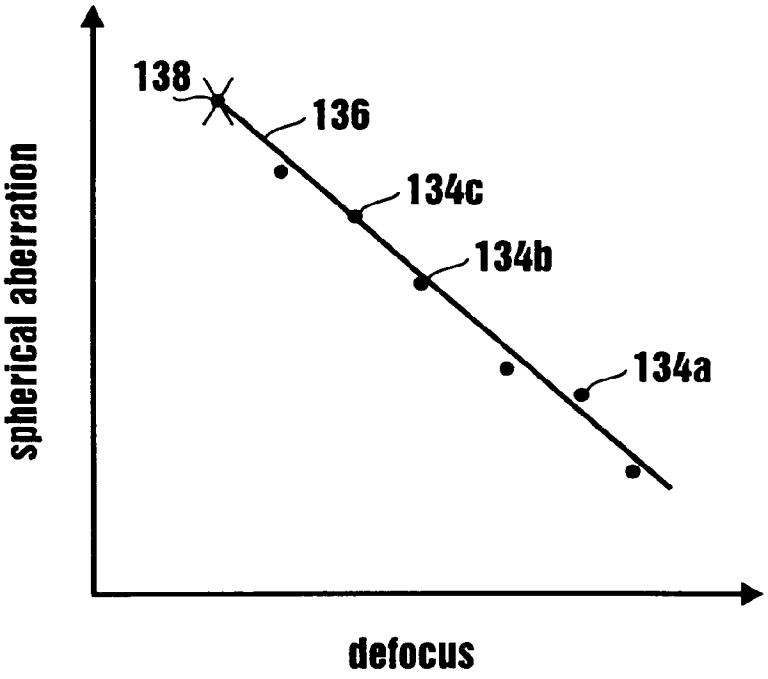



Figure 5

# METHOD AND APPARATUS FOR MEASUREMENT OF THE REFRACTIVE PROPERTIES OF THE HUMAN EYE

## FIELD OF THE INVENTION

The present invention pertains generally to methods and apparatus for performing diagnostic evaluations of the refractive properties of a human eye. More particularly, the present invention pertains to the compilation of information and measurements that are useful for determining and selecting the procedures that are necessary and appropriate for vision correction. The present invention is particularly, but not exclusively, useful for creating acuity and topographical maps of the refractive power of the human eye which can be used for the prescription of corrective elements, such as contact lenses and glasses, or for planning the conduct of refractive surgery.

## BACKGROUND OF THE INVENTION

In the perfect eye, an incoming beam of light is focused through the cornea and through the crystalline lens in a way which causes all of the light from a point source to converge at the same spot on the retina of the eye. This convergence occurs because all of the optical path lengths, for all light in the beam, are equal to each other. Stated differently, in the perfect eye, the time for all light to transit through the eye will be the same regardless of the particular path that is taken by the light.

Not all eyes, however, are perfect. The consequences of this are that light path lengths through the eye become distorted and are not all equal to each other. Thus, light from a point source that transits an imperfect eye will not necessarily be focused on the retina, or to the same spot on the retina.

As light enters and passes through an eye it is refracted at the anterior surface of the cornea, at the posterior surface of the cornea, and at the surfaces of the crystalline lens. It is after all of these refractions have occurred that the light finally reaches the retina. As indicated above, in the case of the perfect eye, all of these refractions result in no overall change in the optical path lengths of light in the incoming beam. Therefore, any deviations which result in unequal changes in these optical path lengths are indicative of imperfections in the eye which may need to be corrected.

In general, vision difficulties in the human eye can be characterized by the changes and differences in optical path lengths that occur as light transits through the eye. These difficulties are not uncommon. Indeed, nearly one half of the world's population suffers from imperfect visual perception. For example, many people are near-sighted because their eyeballs are "too long" (myopia). As a result, the sharp image of an object is generated not on the retina, but in front of or before the retina. Therefore, for a myopic person a distant scene appears to be more or less blurred. On the other hand, hyperopia is a condition wherein the error of refraction causes rays of light entering the eye parallel to the optic axis to be brought to a focus behind the retina. This happens because the eyeball is "too short" from front to back. This condition is commonly referred to as far-sightedness. Unlike the myopic person, a hyperopic, or far-sighted, person will see a near scene as being more or less blurred.

Another refractive malady is astigmatism. Astigmatism, however, is different than either myopia or hyperopia in that it results from an unequal curvature of the refractive surfaces of the eye. With astigmatism, a ray of light is not sharply focused on the retina but is spread over a more or less diffuse area. Further, there are even higher order refractive maladies of interest for vision correction which include coma and spherical aberration. More specifically, coma is an aberration in a lens or lens system whereby an off-axis point object is imaged as a small pear-shaped blob. Coma is caused when the power of the zones of the lens varies with distance of the zone from the axis. Spherical aberration, on the other hand, results from loss of definition of images that are formed by optical systems, such as an eye. Such aberrations arise from the geometry of a spherical surface.



In the past, simple refractive errors of the human eye (myopia, hyperopia and astigmatism) have been corrected conventionally with glasses, dating back to the year 1750. More recently, contact lenses, which were invented about 50 years ago, have been useful for correcting these same more simple refractive errors. Further, refractive laser surgery using Excimer UV-lasers is receiving increased popularity. Thus far, however, all of these techniques for correcting optical impairments of the eye have been limited to the correction of errors from near-sightedness (myopia) or far-sightedness (hyperopia), and to the cylindrical refractive errors, the so-called astigmatism.

As noted above, vision and its refractive errors can be quite complex. Similar to every other optical system, in addition to the simple refractive errors, the human eye also shows higher order refractive errors ("aberrations") such as coma and spherical aberration mentioned above. In all cases, aberrations result when an ideally flat 'wavefront' (i.e. a condition wherein all optical path lengths are equal) is distorted by a real-world optical system. In some cases, these distortions occur in a very complex way. In the trivial case, simple distortions like nearsightedness and far-sightedness would result in an uncomplicated bowl-like symmetrical distortion. With higher order aberrations, however, the result is a complex non-symmetrical distortion of the originally flat wavefront. It is these non-symmetrical distortions which are unique for every optical system, including every single person's eye, and which lead to blurred optical imaging of viewed scenes.

It happens that refractive errors (aberrations or distortions) are stronger when light not only passes through the center of an optical system, but also through the outer regions of the system. Specifically, these aberrations are more pronounced under critical lighting conditions (e.g., twilight). For example, it is well known that people have a comparably small pupil in bright daylight. As the light level decreases, however, the pupil becomes dilated in order to let more light pass through to the retina. With dilation, in addition to passing through the center of the eye light rays will also pass through the outer region of the eye (e.g. the optical system), where the optical quality is low. Thus, even persons with normal 20/20 vision have decreased visual acuity under critical light conditions due to increased higher order aberrations.

A typical approach for improving the vision of a patient has been to first obtain measurements of the eye which relate to the topography of the anterior surface of the cornea. Specifically, such measurements are made to determine the Zernike polynomials. The Zernike polynomials are then used to mathematically describe and to model the anterior surface of the cornea. In accordance with this practice, depending on the order of the Zernike polynomial a certain refractive condition of the eye can be described. For example, the first order terms of the Zernike polynomials describe the tilt of a wavefront while second order terms describe myopia, hyperopia and astigmatism. Third order terms then describe coma and fourth order terms describe i.e. spherical aberration.

Until now, the complex aberrations of the human eye involving coma and spherical aberration could not be measured and, therefore, they could not be corrected. Further, even today, the measurement of the 'standard' so-called simple refractive errors is still not fully objective. In fact, presently the patient's vision is usually categorized using an autorefractor for measuring near-sightedness, far-sightedness, and astigmatism. In the process, cooperation of the patient is crucial for obtaining even rough realistic results with these systems. Still, after this rough initial measurement, the optometrist has to use correction lenses in a subjective procedure to find the corrective strength that is best suited for the patient. To a great extent, these limitations have been caused by an inability to determine a topography for the posterior surface of the eye in addition to determining the topography of the anterior surface. Further, there has been little attention given to the peripheral areas of the cornea where spherical aberrations become more prominent as the pupil of the eye dilates. In order to overcome these deficiencies, it is necessary to evaluate new ways and methods for measuring the refractive characteristics of the cornea.

Heretofore, it has been a common practice to analyze and describe light beams in terms of wavefronts and aberrations of a wavefront. In this regard, the Zernike polynomials have been helpful. A light beam, however, can be conceptualized in a different way; other than as a wavefront. It can also be thought of in terms of a plurality of individual beams, each of which has its own optical path length. Specifically, by way of comparison, at any particular point in time a wavefront can be thought of as being the temporal lengths of the various optical paths that have been traveled by individual light beams from the origin or source of the light. Thus, a light beam with a flat or planar wavefront is equivalent to a light beam wherein all light in the beam has traveled on optical paths that have the same temporal length. A wavefront can be distorted by imperfections in the eye and result in so-called wave aberrations. In terms of optical path lengths, these same aberrations can be thought of as resulting from differences in the optical path lengths of individual beams which are caused by undesirable refractions of light as it passes through the eye.

As discussed above, until now vision correction has been primarily concerned with reshaping the cornea using data that is collected about the topography of the anterior surface of the eye. A good example of technology that is useful for this purpose is provided in U.S. Pat. No. 5,062,702 which issued to Bille for an invention entitled "Device for Mapping Corneal Topography." The posterior surface of the eye, however, also affects the refraction of light as it passes through the eye. Thus, additional information about the thickness of the cornea is necessary for more precise refractive corrections. To this end, a map of the posterior surface of the cornea would undoubtedly be useful. Further, while gross approximations of the lower order visual aberrations using the Zernike polynomials may be useful for limited purposes, the superficial models provided by the Zernike polynomials become quite cumbersome and less precise when higher order aberrations are concerned.

In light of the above, it is an object of the present invention to provide a method and apparatus for measuring the refractive properties of the human eye that are capable of creating a topographical map of the anterior surface of the eye and an acuity map of the refractive power of the entire human eye which are useful either for the prescription of corrective elements or to plan for surgery. Another object of the present invention is to provide a method and apparatus for measuring the refractive properties of the human eye by considering the net effect on individual beams as they each pass through the eye. Yet another object of the present invention is to provide a method and apparatus for measuring the refractive properties of the human eye which, in addition to myopia, hyperopia and astigmatism, can also be used to determine higher order refractive error (aberrations) such as coma and spherical aberration. Still another object of the present invention is to provide a method and apparatus for measuring the refractive properties of the human eye which are effectively easy to use, relatively simple to operate and implement, and comparatively cost effective.

## SUMMARY OF THE PREFERRED EMBODIMENTS

In accordance with the present invention, a system for measuring the refractive properties of the eye includes an optical subsystem for precisely determining the position of the eye. More specifically, this subsystem includes a pupil camera for establishing the general x-y position of the eye, and a confocal detector for precisely establishing the z position of the eye.

Once the position of the eye has been determined and stabilized, a light source is activated to direct a light beam toward the eye for reflection from the anterior surface of the eye. A lenslet is also positioned in the system to separate the light that is reflected from the anterior surface of the eye into a plurality of individual beams which, depending on the topography of the cornea's anterior surface, will each have their own optical path length. These optical path lengths may, or may not, be equal to each other. The individual beams of this plurality are then directed toward a sensor, which cooperates with a computer to create a digital topographical map of the cornea. This topographical map is thus based on the optical path lengths of the individual light beams that are reflected from the anterior surface of the cornea. For purposes of the present invention, light in the beam that is to be reflected from the anterior surface of the cornea will, preferably, have a wavelength of approximately 840 nm and be focused to the center of curvature of the cornea.

Another light source is positioned in the system to direct a light beam toward the eye for reflection from the retina of the eye. Preferably, the light in this beam is collimated as it arrives at the eye, and the beam has a diameter less than about 2 mm as it passes through the pupil toward the retina. The light that is then reflected from the retina fills the pupil and is directed toward a lenslet where, like the light reflected from the cornea, it is separated into a plurality of individual beams. Depending on the particular refractions which have occurred as these individual beams traverse the eye, they may or may not be equal to each other. The individual beams of this plurality are then directed toward a sensor. Like the sensor associated with the individual beams reflected from the cornea, this sensor also cooperates with the computer. The individual beams of light reflected from the retina, however, are used to create a digital acuity map of the entire eye. The acuity map that is thus created, is based on the optical path lengths of the individual light beams that are reflected from the retina. For purposes of the present invention, light in the beam that is to be reflected from the retina will, preferably, have a wavelength of approximately 780 nm.

As intended for the present invention, in order to determine the topography of the posterior surface of the cornea, the refractive effects of the crystalline lens in the eye must

be accounted for. This can be accomplished in either of two ways. For the myopic eye, the patient is required to concentrate on a point at infinity. The crystalline lens in the patient's eye will then be relaxed and, therefore, contribute only its basic relaxed state refraction. On the other hand, for the hyperopic eye, or the infant eye, several successive measurements from successive pluralities of the individual beams that are reflected from the retina are required. It happens, that these measurements, taken in defocused conditions of the lens, are collectively proportional to the relaxed condition. Thus, by using curve fitting techniques, the relaxed condition for the lens can be extrapolated from the data. A determination of the topography of the posterior surface of the cornea is then made by subtracting the topography of the anterior surface of the eye (the topographical map) from the data taken for determining the refractive characteristics of the entire eye (the acuity map).

## BRIEF DESCRIPTION OF THE DRAWINGS

The novel features of this invention, as well as the invention itself, both as to its structure and its operation, will be best understood from the accompanying drawings, taken in conjunction with the accompanying description, in which similar reference characters refer to similar parts, and in which:

FIG. **1** is a schematic drawing of the system of the present invention;

FIG. **2A** is a side cross sectional view of an eye shown in relationship with an incoming planar wavefront of light;

FIG. **2B** is a side cross sectional view of an eye shown in relationship with an outgoing wavefront of light that has been distorted by the refractive properties of the eye;

FIG. **3** is a side cross sectional view of an eye shown with light reflecting from the retina;

FIG. **4** is a side cross sectional view of an eye shown with light reflecting from the anterior surface of the cornea; and

FIG. **5** is a graph showing the relationship between defocus conditions of an eye and the spherical aberration exhibited by the lens of the eye in a respective defocused condition.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring initially to FIG. **1**, a system for determining the refractive properties of the human eye is shown in a schematic drawing and is generally designated **10**. As impliedly indicated in FIG. **1**, use of the system **10** is intended to provide diagnostic information about the eye **12**. In order to do this, the system **10** employs four different light sources and uses four different wavelengths, all for different purposes. More specifically, the system **10** includes a light source **14** which is preferably a laser diode that generates a light beam having a wavelength of approximately eight hundred and forty nanometers (840 nm). Another light source **16** is provided which is preferably a laser diode that will generate a light beam having a wavelength of approximately seven hundred and eighty nanometers (780 nm). There is still another light source **18** which is preferably a laser diode that will generate a light beam having a wavelength of approximately nine hundred and thirty nanometers (930 nm). Finally, there is an illuminator **20** which can include a plurality of infrared diodes that will collectively generate a light beam having a wavelength of approximately nine hundred and eighty nanometers (980).

As intended for the system **10** of the present invention, a computer **22** is used to evaluate the light that is emitted from each of the above mentioned light sources **14**, **16** and **18**, and from the illuminator **20**. More specifically, this evaluation is conducted by the computer **22** after the light from its respective source has been directed toward, and reflected in some way from the eye **12**. With the objective of considering the collective effect of all light that is reflected in the system **10** from the eye **12**, it is perhaps best to first consider each source of light individually, and to track light from that particular source as it passes through the system **10**.

Beginning at the light source **14**, as stated above, a light beam **24** having a wavelength of approximately eight hundred and forty nanometers (840 nm) is generated. Further, the light source **14** directs this beam **24** toward a dichroic beam splitter **26** which will pass 840 nm light, but which will otherwise reflect light that is substantially below 840 nm (e.g. 780 nm). After passing through the beam splitter **26**, the light beam **24** is then reflected by a polarizing beam splitter **28** for further transmission by a beam expander that includes the lenses **30** and **32**. The light beam **24** then passes through a dichroic beam splitter **34** which, like the beam splitter **26**, will pass 840 nm light but reflect 780 nm light. After passing through the beam splitter **34**, the light beam **24** is expanded by the collective effect of lenses **36**, **38** and **40**, and it passes through the dichroic beam splitter **42** toward the dichroic beam splitter **44**. For purposes of the present invention, the dichroic beam splitter **42** will pass light having wavelengths below about 900 nm and will reflect light having wavelengths above about 900 nm. On the other hand, the dichroic beam splitter **44** will pass light having wavelengths above about 830 nm and will reflect light having wavelengths below about 830 nm. Thus, the light beam **24** will pass through both of the beam splitters **42** and **44**. Upon passing through the beam splitter **44**, the light in light beam **24** passes through a quarter wave plate **46** where it is rotated about forty five degrees (45°). The light beam **24** is then focused by a moveable lens **48** onto the eye **12**.

In reflection from the eye **12**, the light beam **24** passes back through the quarter wave plate **46** where it is again rotated an additional forty five degrees (45°). Thus, it is now rotated about ninety degrees (90°) relative to the light in light beam **24** as it was being initially emitted from light source **14**. Further, the reflected light beam **24** passes back through the beam splitters **44**, **42**, and **34**. Due to its dual rotation by the quarter wave plate **46**, however, light beam **24** will not be reflected by the polarizing beam splitter **28**. Instead, the polarizing beam splitter **28** will pass the light beam **24** toward a lenslet array **50** where the beam **24** is separated into a plurality of individual beams. These individual beams are all mutually parallel, and from the lenslet array **50** they are directed toward a dichroic beam splitter **52** which, like beam splitters **26** and **34**, will pass light having a wavelength of 840 nm. After passing through the beam splitter **52**, the individual beams which now comprise the beam **24** are received by an area sensitive detector **54** and are then passed as a plurality of respective signals via the line **56** to the computer **22**. Preferably, the area sensitive detector **54** is a CCD of a type well known in the pertinent art.

In a manner similar to that set forth above for describing the path taken by light beam **24**, consider now the light beam **58** that is generated by the light source **16**. As indicated above, the light beam **58** will have a wavelength that is about 780 nm. Therefore, the light beam **58** will be reflected by the dichroic beam splitter **26** and the polarizing beam splitter **28**. Unlike the light beam **24**, however, the light beam **58** will be reflected by the beam splitter **34** and directed toward the lens **60** and pinhole **62**. A dichroic beam splitter **64** is then provided to direct the light beam **58** through a lens **66**,

toward a turning mirror **68**, and through a focusing lens **70**. It can be noted that for purposes of the present invention, the beam splitter **64** needs to be able to pass visible light below a wavelength of 780 nm (i.e. light in the wavelength range of 780–380 nm). After being reflected by the beam splitter **44**, the light beam **58** is rotated by quarter wave plate **46** and directed by focusing mirror **48** toward the eye **12**. Importantly, as light beam **58** is directed toward the eye **12**, the light in light beam **58** will be substantially collimated and have a beam diameter that is approximately two millimeters (2 mm).

When reflected from the eye **12**, the light beam **58** will again be rotated by the quarter wave plate **46**. Thus, it will be reflected by the beam splitter **44**, turned by turning mirror **68** and reflected by the beam splitters **64** and **34**. Like the light beam **24**, light beam **58** will be passed by the polarizing beam splitter **28** toward the lenslet array **50**. Also like the light beam **24**, the light beam **58** will be separated into a plurality of individual light beams by the lenslet array **50**. The plurality of individual light beams which now comprise the light beam **58** are reflected by the beam splitter **52** and are directed toward an area sensitive detector **72** where the individual beams are converted into respective signals for further transmission via line **74** to the computer **22**.

The light source **18**, as mentioned above, will generate a light beam **75** which has a wavelength of approximately 930 nm. As shown in FIG. **1**, this light beam **75** will pass through a polarizing beam splitter **76**, a lens **78** and a dichroic beam splitter **80**, before being reflected toward the eye **12** by the beam splitter **42**. Importantly, as the beam **75** is directed toward the eye **12** it will pass through and be rotated by the quarter wave plate **46**. The light beam **75** is then reflected from the eye **12**, and light in the reflected light beam **75** will again pass through, and be rotated by, the quarter wave plate **46**. At the beam splitter **42**, the light beam **75** will be directed back toward the polarizing beam splitter **76**. This time, however, light beam **75** does not pass through the beam splitter **76**. Instead, due to its rotations by the quarter wave plate **46**, the light beam **75** is reflected by the polarizing beam splitter **76** through a pinhole **81** toward the confocal detector **82**. A signal that is generated by the confocal detector **82** in response to its reception of the light beam **75** is then passed via line **84** to the computer **22**.

As indicated above, the illuminator **20** generates a light beam **86** which has a wavelength of about 980 nm. For the present invention, the illuminator **20** can include either a plurality of separate infrared diodes, or it can be configured as a ring. In either case, as shown in FIG. **1**, the resultant light beam **86** is pointed directly at the eye **12**. Upon reflection of the light beam **86** from the eye **12**, FIG. **1** indicates that the beam passes through the beam splitter **44** but is reflected by both the beam splitter **42** and the beam splitter **80**. Specifically, insofar as the beam splitter **80** is concerned, it will reflect light such as the light in light beam **86** which has a wavelength greater than about 950 nm. Accordingly, the light in light beam **86** which has been reflected from the eye **12** will be received by a pupil camera **88** and a responsive signal generated by the pupil camera **88** will be sent via line **90** to the computer **22**.

FIG. **1** also shows that the computer **22** is connected via a line **92** with the lens **70**. With this connection the computer **22** is able to adjust the focus of light beam **58** that is provided by lens **70**. Further, FIG. **1** shows that the computer **22** is connected via a line **94** with the lens **48**. With this connection the computer **22** is able to adjust the focus of light beams **24** and **75**. Also, FIG. **1** shows that the computer **22** can include a frame grabber **96** which will provide visual displays of the signals that are received from the area sensitive detectors **54** and **72**, as well as the signals that are received from the confocal detector **82** and the pupil camera **88**.

## OPERATION

In overview, for the purposes of the operation of system **10**, light can be conceptualized in either of two ways. Firstly, light can be thought of in terms of wavefronts. Secondly, light can be thought of as being a collective bundle of many different individually separate beams. These two concepts, of course, must be related to each other if they are to describe the same light beam. Accordingly, in order to reconcile one concept with the other, a wavefront can be thought of as being a spatial representation of the optical path lengths that have been traveled from a common origin (light source) by all of the different individual beams, at any given point in time. Thus, it is the case with unrefracted light which has traveled from a light source in the direction of arrow **98**, as shown in FIG. **2A**, that the light will exhibit a planar wavefront **100**. Stated differently, the optical path length of an individual beam in this light that has traveled from the source to a position **102** in the wavefront **100** will have the same length as the optical path length of an individual beam that has traveled from the source to a position **104** in the wavefront **100**. As the light in wavefront **100** passes through the eye **12**, however, the individual light beams will be refracted differently.

Anatomically, it is necessary for light entering the eye **12** to pass through the cornea **106** and the crystalline lens **108** of the eye **12** before coming into contact with the retina **110**. It is known that our visual perception is dependent on how this light comes into contact with the retina **110** and, it is known, of course, that in accordance with Snell's law this light will be refracted by the cornea **106** and the lens **108** as it passes into the eye **12**. Further, whatever light there is that is reflected from the retina **110** to pass back through the eye **12** and away from the eye **12** will also be refracted by the lens **108** and the cornea **106**. The result of all this refraction may likely be a distorted wavefront **112** which is traveling away from the eye **12** in the direction of arrow **114**, such as shown in FIG. **2B**. By comparing FIG. **2A** with FIG. **2B** it will be noted that, due to refractions caused by the eye **12**, the optical path length of the individual beam which has traveled from position **102** in the planar wavefront **100** to the position **102'** in the distorted wavefront **112** will be different from the optical path length of the individual beam which has traveled from position **104** to position **104'**. As appreciated by the present invention, the differences in these optical path lengths is indicative of the respective refractions that were experienced by the individual beams as they transited the eye **12**.

In the operation of the system **10** of the present invention, it may first be necessary to calibrate the system **10**. This can be done by replacing the eye **12** with a flat mirror (not shown). Light can then be sequentially passed through the system **10** from the light sources **14**, **16** and **18** and from the illuminator **20** for reflection back through the system **10** from the flat mirror. In this way, signals can be generated from the area sensitive detectors **54** and **72**, from the confocal detector **82** and from the pupil camera **88**. The signals which are thus generated will be indicative of inherent optical aberrations in the system **10** and can subsequently be used to compensate actual signals generated by the eye **12**.

Once the system **10** has been calibrated, it is desirable to determine an exact spatial position for the eye **12** in x-y-z.

This is done by using the confocal detector **82** and the pupil camera **88**. Specifically, in order to establish a "z" position for the eye **12**, the light source **18** is activated to generate the light beam **75**. For the present invention, the light beam **75** is focused by the lens **48** to obtain a specular reflection of the light beam **75** from the apex **114** of the cornea **106** (see FIG. **3**). Depending on the position of the lens **48**, which is sensed by the computer **22**, when this specular reflection is obtained the position of the eye **12**, and more specifically the position of the apex **114** of eye **12**, in a "z" direction along the visual axis of the eye **12** is established. In order to establish an "x-y" position for the eye **12**, the illuminator **20** needs to be activated. Specifically, using intensity differences in the reflection of light beam **86** from the eye **12**, as sensed by the pupil camera **88** and ultimately by the computer **22**, the "x-y" position of the eye **12** is established. For the present invention, the intensity differences used for this measurement are cause by the contrast between the iris **116** and the lens **108** at the periphery of the pupil **118**. While the "z" position has been considered as having been taken first in this discussion, it will be appreciated by the skilled artisan that the "x-y" determination may, in fact, be made first.

Once the system **10** is calibrated and the position of the eye **12** in "x-y" and "z" has been established, refractive measurements of the eye can be made. In light of the above discussion, it will be noted that these measurements need to be considered with knowledge of the length of the eye **12**, and with an understanding that the crystalline lens **108** is in its basic relaxed state of refraction. Insofar as a measurement for the length of the eye **12** is concerned, this can be done by activating the light source **114** while the eye **12** is focused to a point **124** at infinity (see FIG. **1**). The focusing lens **70** is then moved as required by the computer **22**, or manually by the operator of system **10**, until the light beam **24** from light source **114** is focused onto the retina **110**. By using the position of the lens **70** for this focal condition, the computer **22** is able to establish a position of the retina **110**. Then, by knowing the position of the retina **110**, and by knowing the location of the apex **114** that was obtained from earlier measurements of the "z" position of the eye **12**, the length of the eye **12** can be determined. This measurement, of course, will also determine whether the eye **12** is myopic or hyperopic.

Referring to FIG. **3**, it will be appreciated that if the light beam **24** is focused by the eye **12**, without any correction, to a location **120** in front of the retina **110**, and movement of the lens **70** is necessary to move the focal point of light beam **24** from the location **120** backward toward the retina **110**, the eye **12** is myopic. On the other hand, if the light beam **24** is focused by the eye to a location **122** behind the retina **110**, again without any correction, and movement of the lens **70** is necessary to move the focal point of light beam **24** from the location **122** forward toward the retina **110**, the eye **12** is hyperopic. The determination of whether the eye **12** is myopic or hyperopic is important, not only in its own right, but also for subsequent refractive measurements. Importantly, while the myopic eye **12** focuses on the point **124** at infinity, it can be taken that the crystalline lens **70** will be in its basic relaxed state of refraction. On the other hand, as mentioned above and as explained in more detail below, when the eye **12** is hyperopic of is the eye of an infant several successive measurements need to be taken.

The initial measurement for the general topography of the cornea **106** of the eye **12** is made using the light beam **24** that is generated by light source **14**, and which has a wavelength of 840 nm. By cross referencing FIG. **1** with FIG. **4** it will be seen that the light beam **24** is focused by the lens **48** toward the center of curvature **126** of the cornea **106**. In doing this, the light beam **24** covers a distance **128** on the anterior surface **130** of the cornea **106** that is equal to about seven millimeters (7 mm). Importantly, some of the light in the light beam **24** will be reflected by the anterior surface **130** of the cornea **106** and will be directed back through the system **10** to the lenslet **50** where it is separated into a plurality of individual beams. These individual beams in the reflected light of light beam **24** are then detected by the area sensitive detector **54** which generates signals that are sent via line **56** to the computer **22**. Using these particular signals, the computer **22** is able to create a topography map of the anterior surface **130** of the cornea **106**.

For the present invention, refractive measurements of the entire eye **12** are made using the light beam **58** that is generated by light source **16**. The shorter wavelength of 780 nm is selected for the light beam **58** because it is near the visible range and it will, therefore, more easily travel through the eye **12** than will the longer wavelength light beams used for other purposes in the system **10**. It is an important consideration that the light beam **58** have a relatively small cross section as it initially enters the eye **12**. This is so in order to minimize light refractions that are caused as the light beam **58** travels through the eye **12** toward the retina **110**. For the present invention, the light beam **58** is, preferably, confined to about two millimeters (2 mm) diameter. Also, the light beam **58** is adjusted by the optics along its beam path so that as the light beam **58** leaves the lens **48** and travels toward the eye **12**, it is substantially collimated when it arrives at the anterior surface **130** of the cornea **106**.

Returning to FIG. **3**, it will be seen that the light in light beam **58** will be reflected from the retina **110** as a light beam **58'** which fills the pupil **118**. This reflected light beam **58'** then is passed back through the system **10** to the lenslet **50** where, like the beam **24**, it is separated into a plurality of individual beams. Also like the individual beams of light beam **24**, the individual beams of light beam **58** are passed to an area sensitive detector **72** where signals are generated for transmission via line **74** to the computer **22**. More specifically, the individual beams of light beam **58** are collectively used to generate an acuity map of the eye **12** which is indicative of the refractions experienced by light as it passes through the eye **12**.

For a myopic eye **12**, which will remain in its basic relaxed state of refraction while focused to infinity on the point **124** (see FIG. **1**), a topography for the posterior surface **132** of the cornea **106** can be determined using computer **22**. Basically, this is done by subtracting the topography map data for the anterior surface **130** of the eye **12**, and the basic relaxed state refraction for the crystalline lens **108**, from the acuity map of the entire eye **12**. The result is data which can be used directly to determine the topography for the posterior surface **132** of the eye **12**.

For a hyperopic eye **12** or an infant eye **12**, which is not able to establish its basic relaxed state of refraction while focused to infinity on the point **124**, successive measurements need to be taken and the collected data extrapolated to determine the basic relaxed state of refraction. This is possible because it is known that there is a generally linear relationship between each defocus condition of the crystalline lens **108** in the eye **12**, and the corresponding spherical aberrations caused by the lens **108** (see FIG. **5**). Therefore, by taking a series of successive measurements for the acuity map (i.e. using the light beam **58** from light sources **16**) a plurality of data points **134** (of which the data points **134***a*, **134***b* and **134***c* are representative), can be plotted. In FIG. **5**

it is seen that the plots of data points **134** can be used to identify a line **136** and that the point **138** can be extrapolated and be considered equivalent to the conditions extant when the crystalline lens **108** is in its basic relaxed state of refraction. In a manner as disclosed above in consideration of the myopic eye **12**, this data can be used with the topography map of the anterior surface **130** of the cornea **106** and the acuity map of the entire eye **12** to determine a topography map for the posterior surface **132** of the eye **12**. In any case, all of the data collected will give the operating physician a much more detailed measurement of the anatomy of the eye **12** which will be useful for the prescription of corrective elements or for planning the conduct of refractive surgery.

While the particular method and apparatus for measurement of the refractive properties of the human eye as herein shown and disclosed in detail is fully capable of obtaining the objects and providing the advantages herein before stated, it is to be understood that it is merely illustrative of the presently preferred embodiments of the invention and that no limitations are intended to the details of construction or design herein shown other than as described in the appended claims.

What is claimed is:

**1**. A system for measuring optical properties of an eye, the eye having, in order, a cornea, a lens and a retina, said system comprising:

an optical means for directing a first light beam for reflection from the cornea of the eye as a reflected first light beam;

a lenslet array for separating said reflected first light beam into a plurality of first individual beams with each said first individual beam having an optical path length;

an optical means for aiming a second light beam through the cornea, and through the lens, for reflection from the retina of the eye as a reflected second light beam;

a lenslet array for separating said reflected second light beam into a plurality of second individual beams with each said second individual beam having an optical path length;

a computer means for using said plurality of first individual beams to create a topographical map of the cornea indicative of the optical path lengths of said first individual beams, and for using said plurality of second individual beams to create an acuity map of the eye indicative of the optical path lengths of said second individual beams; and

a comparator means operable with said computer means for comparing said topographical map of the cornea with said acuity map of the eye to determine selected optical properties of the eye.

**2**. A system as recited in claim **1** wherein the optical properties of the eye are determined by differences between the optical path lengths of said first individual beams used for creation of said topographical map, and by differences between the optical path lengths of said second individual beams used for creation of said acuity map.

**3**. A system as recited in claim **1** further comprising a lens for focusing said second light beam onto the retina to determine a length for the eye.

**4**. A system as recited in claim **1** wherein said computer means analyzes said topographical map to model an anterior surf ace of the cornea.

**5**. A system as recited in claim **1** wherein said topographical map is indicative of aberrations for the anterior surface of the cornea, wherein said acuity map is indicative of aberrations for the entire eye, and wherein said comparator means separates the aberrations of the anterior surface of the cornea and a predetermined aberration for the lens, from the aberrations of the entire eye to determine aberrations for the posterior surface of the cornea.

**6**. A system as recited in claim **5** wherein said aberrations of the posterior surface of the cornea are used by said computer means to model the posterior surface of the cornea.

**7**. A system as recited in claim **5** wherein the spherical aberration of the lens is determined with the lens in a relaxed state.

**8**. A system as recited in claim **5** wherein the predetermined aberration of the lens is determined by said computer means using a plurality of sequentially created said acuity maps.

**9**. A system as recited in claim **1** further comprising:

an optical means for sending a third light beam for reflection from the apex of the cornea as a reflected third light beam;

an illuminator for generating a pupil reflection from the eye; and

a detector means for analyzing said pupil reflection for determining an "x-y" position for the eye, and for analyzing said reflected third light beam for establishing a "z" position for the eye.

**10**. A system as recited in claim **1** wherein the cornea has a center of curvature and said first light beam is focused onto the center of curvature of the cornea, and wherein said first light beam has a wavelength equal to approximately eight hundred and forty nanometers (840 nm).

**11**. A system as recited in claim **1** wherein the eye has a pupil and said second light beam is aimed toward the eye as collimated light for subsequent focus by the cornea and the lens of the eye toward the retina, and wherein said second light beam has a wavelength equal to approximately seven hundred and eighty nanometers (780 nm), and further wherein said second light beam has a diameter of approximately 2 mm when entering through the pupil and a diameter of approximately 6 mm when emerging from the eye through the pupil after reflection from the retina.

**12**. A system for measuring optical properties of the eye, the eye including a cornea having an anterior surface and a posterior surface, a lens and a retina, said system comprising:

means for creating a topographical map of the anterior surface of the eye, said topographical map including a plurality of individual refractive measurements;

means for generating an acuity map for the entire eye, said acuity map comprising a plurality of individual refractive measurements;

means for determining an aberration for a relaxed lens; and

computer means for comparing said topographical map with said acuity map while compensating for the aberration of the lens to construct a topography for the posterior surface of the eye.

**13**. A system as recited in claim **12** wherein said means for creating a topographical map comprises:

an optical means for directing a first light beam for reflection from the anterior surface of the cornea of the eye as a reflected first light beam; and

a lenslet array for separating said reflected first light beam into a plurality of first individual beams with each said first individual beam having an optical path length and each said optical path length being indicative of a respective said refractive measurement.

14. A system as recited in claim 13 wherein said means for generating an acuity map comprises:

an optical means for aiming a second light beam through the cornea, and through the lens, for reflection from the retina of the eye as a reflected second light beam;

a lenslet array for separating said reflected second light beam into a plurality of second individual beams with each said second individual beam having an optical path length and each said optical path length being indicative of a respective said refractive measurement.

15. A system as recited in claim 14 further comprising:

an optical means for sending a third light beam for reflection from the apex of the cornea as a reflected third light beam;

an illuminator for generating a pupil reflection from the eye; and

a detector means for analyzing said pupil reflection for determining an "x-y" position for the eye, and for analyzing said reflected third light beam for establishing a "z" position for the eye.

16. A system as recited in claim 14 wherein the cornea has a center of curvature and said first light beam is focused onto the center of curvature of the cornea, and wherein said first light beam has a wavelength equal to approximately eight hundred and forty nanometers (840 nm), and further wherein the eye has a pupil and said second light beam is aimed toward the eye as collimated light for subsequent focus by the cornea and the lens of the eye toward the retina, and wherein said second light beam has a wavelength equal to approximately seven hundred and eighty nanometers (780 nm) and has a diameter of approximately 2 mm when entering through the pupil and a diameter of approximately 6 mm when emerging from the eye through the pupil after reflection from the retina.

17. A method for measuring optical properties of the eye, the eye including a cornea having an anterior surface and a posterior surface, a lens and a retina, said method comprising the steps of:

creating a topographical map of the anterior surface of the eye, said topographical map including a plurality of individual refractive measurements;

generating an acuity map for the entire eye, said acuity map comprising a plurality of individual refractive measurements;

determining an aberration for a relaxed lens; and

comparing said topographical map with said acuity map while compensating for the aberration of the lens to construct a topography for the posterior surface of the eye.

18. A method as recited in claim 17 wherein said step of creating a topographical map comprises the steps of:

directing a first light beam for reflection from the anterior surface of the cornea of the eye as a reflected first light beam; and

separating said reflected first light beam into a plurality of first individual beams with each said first individual beam having an optical path length and each said optical path length being indicative of a respective said refractive measurement.

19. A method as recited in claim 18 wherein said step of generating an acuity map comprises the steps of:

aiming a second light beam through the cornea, and through the lens, for reflection from the retina of the eye as a reflected second light beam; and

separating said reflected second light beam into a plurality of second individual beams with each said second individual beam having an optical path length and each said optical path length being indicative of a respective said refractive measurement.

20. A method as recited in claim 17 further comprising the steps of:

sending a third light beam for reflection from the apex of the cornea as a reflected third light beam;

illuminating the eye to generate a pupil reflection from the eye; and

analyzing said pupil reflection for determining an "x-y" position for the eye; and

analyzing said reflected third light beam for establishing a "z" position for the eye.

* * * * *

# APPENDIX E

Reprinted from *Journal of the Optical Society of America,* Vol. 65(5), pp. 551-558 (May 1975).

# Photon noise and atmospheric noise in active optical systems

Freeman J. Dyson*

*Max-Planck-Institut für Physik und Astrophysik, Föhringer Ring 6, 8 München 40, Germany*
(Received 18 October 1974; revision received 3 December 1974)

A general theoretical analysis is made of the performance of optical systems that are designed to give diffraction-limited images of astronomical objects by compensating effects of atmospheric seeing in real time. The heart of any such system is a feedback algorithm, which expresses the controlled displacements of mirror surfaces as functions of the output of optical sensors. The statistical behavior of the system is calculated, assuming the feedback algorithm to be linear but otherwise unrestricted. The effect of feedback in diminishing atmospheric noise while amplifying photon noise is worked out in detail, taking into account photon–photon correlations. A figure of merit for system performance, namely a statistical average of a positive quadratic function of phase errors over mirror surfaces, is arbitrarily chosen. By use of this figure of merit, the optimum feedback algorithm for any given optical system is explicitly determined. The optimum algorithm is independent of the quadratic function of phase errors that is chosen as the figure of merit. Applications of the general theory to particular optical systems are briefly discussed. In principle, systems optimized in this way should be able to give images of arbitrarily high resolution of astronomical objects brighter than about magnitude 14.

The phrase active optical system in this paper means a system designed to provide high-resolution images of celestial objects, with movable optical surfaces that compensate in real time the effects of fluctuating atmospheric perturbation of the incoming light. Such a system was first proposed by Babcock,[1] and has been recently investigated experimentally by Muller and Buffington,[2] Bridges *et al.*,[3] Hardy *et al.*,[4] Miller *et al.*,[5] Ogrodnik,[6] and Pearson *et al.*[7] Every active optical system consists conceptually of six parts: A, one or more light-collecting mirrors, either in the form of a conventional telescope, an interferometer, or an array of interferometers; B, a secondary flexible mirror, in which each small patch of surface can be moved independently to introduce a controlled variation of optical path length of the light reflected from that patch; C, a set of electromechanical relays that control the movements of the patches of surface in B; D, a television monitor or array of photon detectors that record the places and times of arrival of photons that pass through the system; E, a computer that accepts input from D and which, on the basis of this information, feeds output to C, so that the movement of the mirror surface B is responsive to the optical performance of the system at previous times; F, a computer program or algorithm that instructs the computer E how to process the input and compute the output. The essential task that the computer program has to perform is to tell the flexible mirror what to do in order to improve the optical quality of the image, using for this purpose the information provided by the image itself.

The five components A–E of the system are hardware items, which have been extensively discussed by the experimenters.[2–7] In this paper, we take the hardware components as given and concentrate our attention on the design of the program F. To design an effective computer algorithm is equivalent to solving a prediction problem, namely to predict the behavior of atmospheric perturbations at time $t$ when we know only the optical response of the system to perturbations that occur at times earlier than $t$. The possibility of an effective prediction algorithm is, in principle, limited by two sources of noise, first the statistical fluctuations of the atmospheric perturbations (atmospheric noise), and second the statistical fluctuations of the output of the optical sensors (photon noise). The feedback loop, from optical sensor to computer to servomechanism to flexible mirror to optical sensor, has the primary function of diminishing the atmospheric noise, but the photon noise is necessarily amplified by the feedback as the atmospheric noise is attenuated. It is to be expected that an optimum strength of feedback will be obtained when the two sources of noise are made approximately equal. The exact optimization of this tradeoff between atmospheric and photon noise will depend in a sensitive way upon the frequency spectrum of the atmospheric perturbations.

This paper consists of three parts. In Sec. I, we make no restriction on the nature of the feedback algorithm except that it is to be linear and causal. That is to say, we let the computer outputs (mirror-patch displacements at time $t$) be arbitrary linear functions of the computer inputs (optical sensor responses) at times previous to $t$. We then compute the statistical behavior of the whole system responding to atmospheric perturbations of given statistical characteristics. The crucial feature of this analysis is an exact treatment of photon–photon correlations. Photon–photon correlations arise because the probability of detection of a photon in a sensor at time $t_2$ is significantly changed by the mirror displacement induced by the feedback response to a single photon detected at an earlier time $t_1$. The degradation of system performance by photon noise depends in an essential way on these photon–photon correlations. The end result of Sec. I is a compact formula (42) that exhibits the dependence of the system performance upon the atmospheric-fluctuation spectrum, the optical characteristics of the system, and the feedback algorithm.

In Sec. II, we use the results of Sec. I to solve exactly

the problem of optimizing the feedback algorithm for a system whose hardware components are given. Here, optimization has to be defined in some arbitrary fashion. We choose to define optimization as the minimization of a mean-square deviation (averaged over both atmospheric and photon fluctuations) of optical phase errors over the mirror surfaces. Fortunately, it turns out that the optimum feedback algorithm so defined is independent of which particular mean-square average deviation we choose to minimize. The optimized algorithm may therefore be expected to be useful even under conditions in which our arbitrarily chosen criterion of optimization is inappropriate.

The characteristics of the optimum algorithm are determined in two steps. First, we determine the space dependence of the feedback by means of a duality theorem (Theorem 1), which states that the feedback from optical sensor $j$ to mirror patch $k$ should be proportional to the optical response of the sensor $j$ to a displacement of the patch $k$. Second, we determine the time dependence of the feedback by a nonlinear integral equation (Theorem 2), which relates the feedback directly to the space–time correlation structure of the atmospheric fluctuations. The two theorems together determine the optimum algorithm uniquely. It turns out that the nonlinear integral equation (85) is identical to the Gelfand–Levitan equation,[8] which enables an unknown scattering potential to be reconstructed from the scattering phase shifts that it induces upon waves passing through it. It is perhaps not accidental that the same equation arises in the present context, because we are dealing with the reconstruction of an unknown atmospheric-scattering medium from observations of phase shifts in the light that comes through it. Nevertheless, the analogy is far from exact. Gelfand and Levitan were not discussing an optimization problem, and their space coordinate plays the role that in our analysis is played by the time coordinate. The fact that precisely the same equation appears in these two different physical contexts indicates a connection whose exact logical basis remains to be explored.

In Sec. III, we indicate how the general results of Secs. I and II are to be applied to particular optical systems, and make a rough estimate of the brightness of objects for which effective compensation of atmospheric perturbations might be achieved. Section III is purposely made brief, because an informed discussion of the practical aspects of active optical systems lies beyond the competence of the author.

## I. PHOTON-PHOTON CORRELATIONS

First, we introduce notations designed to characterize an active optical system in a general abstract fashion. The advantage of these abstract notations is that the equations become so compact that their structure is easily grasped.

The patches of movable mirror surface are indexed by the letter $k$. It is convenient to think of $k$ as a discrete variable, but the theory applies unchanged with a continuously variable $k$ if the movable mirror is regarded as a continuously deformable surface. The photon-detector channels are indexed by the letter $j$. Here $j$ must in reality be a discrete variable, but it may be convenient to idealize the detector system by imagining $j$ to indicate the precise position of a photon arriving at the image plane, or the precise wave number of a photon passing through a spectrometer, in which case $j$ would take a continuous range of values. When we refer to the state of mirror patch $k$ at the time $t_k$, we use the letter $y$ instead of the pair $(k, t_k)$. Thus, the imposed displacement of the mirror surface is defined by a function

$$\tfrac{1}{2}\, b = \tfrac{1}{2}\, b(y) = \tfrac{1}{2}\, b(k, t_k) \tag{1}$$

whose values are controlled by the computer output. Similarly, the letter $x$ denotes a detector channel $j$, together with a time $t_j$, so that

$$P = P(x) = P(j, t_j) \tag{2}$$

denotes the probability per unit time that a photon will arrive at channel $j$ at time $t_j$. Because of the quantum nature of light, the actual photon count in channel $j$ is not $P$ but

$$I = I(x) = I(j, t_j). \tag{3}$$

We have

$$P = \langle I \rangle_p, \tag{4}$$

where the bracket $\langle\rangle_p$ denotes an average over photon statistics, and $P$ is the irradiance calculated from classical electromagnetic theory, given any particular configuration of the atmosphere and the mirror displacements.

We suppose that the unknown atmospheric perturbation whose effect the system is designed to compensate is denoted by a function

$$a = a(y) = a(k, t_k), \tag{5}$$

representing the fluctuation of optical path length of the pencil of light rays that arrives from the object at the mirror patch $k$. The field of view of the system is assumed to be so small that all rays from the object to a given mirror patch pass through the same isoplanatic patch in the sky and are subject to the same atmospheric perturbation. The path-length perturbation $a$ is supposed to be compensated in part by the mirror displacement $\frac{1}{2}b$, so that there is a net path-length error

$$s = a + b = s(k, t_k) \tag{6}$$

in the light waves reflected from mirror patch $k$. The statistical characteristics of the atmosphere are defined by a space–time correlation function

$$\begin{aligned} U &= U(y_1, y_2) = U(k_1, k_2, t_1 - t_2) \\ &= \langle a(k_1, t_1) a(k_2, t_2) \rangle_a, \end{aligned} \tag{7}$$

where the bracket $\langle\rangle_a$ denotes an average over atmospheric statistics. The numerical values of $U$ will depend on seeing conditions, wind speeds and other meteorological variables that are not explicitly represented in Eq. (7).

When a given astronomical object is observed by the system, classical electromagnetic theory predicts a distribution of irradiance $P$ in the various detector chan-

nels, with $P$ depending in a calculable way on the path-length errors $s$ given by Eq. (6). We assume that the feedback system is successful in keeping the errors $s$ small compared with an optical wavelength, so that the dependence of $P$ on $s$ is approximately linear. We have then a linear equation,

$$P(j,t) = P_0(j) + \sum_k B_{jk} s(k,t), \tag{8}$$

which we write in operator notation as

$$P = P_0 + Bs. \tag{9}$$

Here $P_0$ is the distribution of irradiance in the detector channels in the absence of atmospheric perturbations, and $B$ is the optical-response kernel

$$B(x,y) = B_{jk}\delta(t_j - t_k) \tag{10}$$

that gives the instantaneous optical response of the detector $j$ to a displacement of the mirror patch $k$. Both $P_0$ and $B$ depend on only the optical characteristics of the object and the observing apparatus. They are independent of the atmosphere and of the feedback system.

The feedback algorithm is assumed to be a linear relation

$$b(k,t) = \sum_j \int_{-\infty}^{t} dt' N_{kj}(t-t') I(j,t'), \tag{11}$$

or in operator notation

$$b = NI, \tag{12}$$

where the feedback kernel $N$ relates the mirror displacement $b$ to the photon count $I$ at earlier times. It is important that $I$, rather than $P$, appears on the right-hand side of Eq. (12), so that photon fluctuations are included in the feedback. The system should be in equilibrium with zero feedback when the irradiance has the unperturbed distribution $P_0$, so that we require the feedback kernel to satisfy the condition

$$NP_0 = 0. \tag{13}$$

It will turn out that Eq. (13) is satisfied automatically by the optimized feedback kernel, which we shall compute in Sec. II.

We have now a complete abstract definition of the system in terms of optical-sensor output $I$, mirror displacement $\frac{1}{2}b$, atmospheric path-length perturbation $a$, atmospheric correlation function $U$, unperturbed irradiance $P_0$, optical response kernel $B$, and feedback kernel $N$. These quantities are related by the closed system of Eqs. (4), (6), (7), (9), (12), and (13). It remains to calculate the performance of the system by solving these equations.

We begin by calculating the one-point functions or averages of quantities linear in the photon counts. These quantities obey linear equations obtained by averaging Eqs. (9) and (12) over the photon statistics,

$$\langle s \rangle_p = a + N(\langle I \rangle_p - P_0), \tag{14}$$

$$\langle I \rangle_p - P_0 = B\langle s \rangle_p. \tag{15}$$

These equations have the solution

$$\langle s \rangle_p = [1 - NB]^{-1} a, \tag{16}$$

$$\langle I \rangle_p = P_0 + [1 - BN]^{-1} Ba. \tag{17}$$

The operators on the right-hand side of Eqs. (16) and (17) can be inverted in a straightforward manner, by using Fourier transforms or otherwise. The equations show that the feedback reduces the mean path-length errors $\langle s \rangle_p$ to a small fraction of the atmospheric perturbations $a$, provided that the product $NB$ is made large and negative. Only the effects of photon noise set an upper limit to the strength of negative feedback that can be usefully applied. Photon-noise effects appear in the two-point functions or averages of quantities bilinear in the photon counts. We now proceed to calculate these.

We use the abbreviated notation

$$s_1 = s(y_1), \quad I_1 = I(x_1) \tag{18}$$

to denote the value of the quantities $s$ and $I$ at places and times labelled by the index 1. The two-point functions are then

$$\langle s_1 s_2 \rangle_p = \langle s(y_1) s(y_2) \rangle_p, \tag{19}$$

and similarly $\langle s_1 I_2 \rangle_p$ and $\langle I_1 I_2 \rangle_p$. Two equations for these functions are obtained immediately by multiplying Eq. (12) by the value of $s$ or $I$ at a second place and taking the expectation value. In this way, we obtain

$$\langle s_1 s_2 \rangle_p = \langle s_1 \rangle_p a_2 + N_2 \langle s_1 I_2 \rangle_p, \tag{20}$$

$$\langle s_1 I_2 \rangle_p = a_1 \langle I_2 \rangle_p + N_1 \langle I_1 I_2 \rangle_p. \tag{21}$$

Here the notation $N_1$ means the kernel $N$ operating on the space–time coordinates labelled by the index 1 only, and similarly for $N_2$. Equations (20) and (21) hold, irrespective of the time sequence of the places 1 and 2. The multiplier function [$s_1$ in Eq. (20), $I_2$ in Eq. (21)] plays only the role of a spectator and does not disturb the operation of the feedback kernel.

The third equation for the two-point functions must be derived from Eq. (9). The situation is here more complicated, because the quantity on the left-hand side of Eq. (9) is the expectation value $\langle I \rangle_p$ and not $I$ itself. If we multiply Eq. (9) by the value of $I$ at a second place, we obtain

$$\langle P_1 I_2 \rangle_p = P_{01} \langle I_2 \rangle_p + B_1 \langle s_1 I_2 \rangle_p. \tag{22}$$

Now, according to Eq. (4)

$$\langle P_1 I_2 \rangle_p = \langle \langle I_1 \rangle_p I_2 \rangle_p, \tag{23}$$

where the inner average is over photon statistics at the time $t_1$ only, whereas the outer average is over photon statistics at all times. It is permissible to write

$$\langle \langle I_1 \rangle_p I_2 \rangle_p = \langle I_1 I_2 \rangle_p \tag{24}$$

if and only if $t_1 > t_2$, so that the photons detected at time $t_1$ cannot have any causal influence upon photons detected at $t_2$. Equation (22) thus becomes

$$\langle I_1 I_2 \rangle_p = P_{01} \langle I_2 \rangle_p + B_1 \langle s_1 I_2 \rangle_p, \tag{25}$$

valid for $t_1 > t_2$ only. The function $\langle I_1 I_2 \rangle_p$ must also contain a singular contribution that arises when the places 1 and 2 coincide in space and time. To evaluate this singular contribution, we consider the quantity

$$n = \int_R I(x)\,dx, \tag{26}$$

where the $x$ integration is over any subset of detector channels and over a very short interval of time $\delta t$. If $\delta t$ is short enough, at most one photon will be detected within the integration volume, and so $n$ is a random variable taking only the values 0 and 1. We therefore have

$$\langle n^2\rangle_p = \langle n\rangle_p, \tag{27}$$

or explicitly

$$\int\int_R dx_1\,dx_2\,\langle I_1 I_2\rangle_p = \int_R dx\,\langle I\rangle_p, \tag{28}$$

to first order in $\delta t$. Equation (28) implies that $\langle I_1 I_2\rangle$ contains the singular contribution

$$\Delta_{12} = \delta_{12} P_1, \tag{29}$$

where

$$\delta_{12} = \delta(j_1, j_2)\delta(t_1 - t_2) \tag{30}$$

is a delta function that identifies the places 1 and 2 in both channel index and time. Equation (25), supplemented by the singular term, becomes

$$\langle I_1 I_2\rangle_p = P_{01}\langle I_2\rangle_p + B_1\langle s_1 I_2\rangle_p + \Delta_{12}, \tag{31}$$

valid now for $t_1 \geq t_2$.

It is convenient to introduce the reduced two-point functions defined by

$$\langle s_1 s_2\rangle_{pr} = \langle s_1 s_2\rangle_p - \langle s_1\rangle_p\langle s_2\rangle_p, \tag{32}$$

and similarly for $\langle s_1 I_2\rangle_{pr}$ and $\langle I_1 I_2\rangle_{pr}$. The reduced functions represent purely the effects of photon–photon correlations. Equations (20), (21), and (31) give three equations for the reduced functions,

$$\langle s_1 s_2\rangle_{pr} = N_2\langle s_1 I_2\rangle_{pr}, \tag{33}$$

$$\langle s_1 I_2\rangle_{pr} = N_1\langle I_1 I_2\rangle_{pr}, \tag{34}$$

$$\langle I_1 I_2\rangle_{pr} = B_1\langle s_1 I_2\rangle_{pr} + \Delta_{12}; \tag{35}$$

the last equation holds for $t_1 \geq t_2$ only. From Eqs. (34) and (35), we deduce

$$(1 - B_1 N_1)\langle I_1 I_2\rangle_{pr} = \Delta_{12}, \quad t_1 \geq t_2. \tag{36}$$

We symmetrize Eq. (36) by operating on both sides with the operator $(1 - B_2 N_2)$. Because $B_2$ is an instantaneous kernel, whereas $N_2$ is causal, the resulting equation still holds for $t_1 \geq t_2$. Also, because $\Delta_{12}$ is instantaneous, the product $B_2 N_2 \Delta_{12}$ is zero for $t_1 \geq t_2$. We have therefore

$$(1 - B_1 N_1)(1 - B_2 N_2)\langle I_1 I_2\rangle_{pr} = \Delta_{12} \tag{37}$$

for $t_1 \geq t_2$. But because Eq. (37) is symmetric, it holds irrespective of the time ordering. The effects of photon–photon correlation are then given explicitly by Eq. (33), (34), and (37) in the form

$$\langle I_1 I_2\rangle_{pr} = (1 - B_1 N_1)^{-1}(1 - B_2 N_2)^{-1}\Delta_{12}, \tag{38}$$

$$\langle s_1 s_2\rangle_{pr} = (1 - N_1 B_1)^{-1}(1 - N_2 B_2)^{-1} N_1 N_2 \Delta_{12}. \tag{39}$$

Going back to the unreduced two-point function, we find from Eqs. (16) and (39)

$$\langle s_1 s_2\rangle_p = (1 - N_1 B_1)^{-1}(1 - N_2 B_2)^{-1}\{a_1 a_2 + N_1 N_2 \Delta_{12}\}. \tag{40}$$

In assessing the performance of the whole system, we are finally interested in the statistical fluctuations of the path-length error $s$ under the combined influence of atmospheric and photon noise. We use the double bracket $\langle\langle\ \rangle\rangle$ to denote an average over both atmospheric and photon statistics. We assume that the feedback algorithm is chosen to make the mean error

$$\langle\langle s\rangle\rangle = (1 - NB)^{-1}\langle a\rangle_a = 0, \tag{41}$$

which condition will usually be satisfied trivially, because the mean atmospheric perturbation $\langle a\rangle_a$ will normally be zero. Then the double average applied to Eq. (40) gives, by virtue of Eqs. (7), (9), and (29),

$$\langle\langle s_1 s_2\rangle\rangle = (1 - N_1 B_1)^{-1}(1 - N_2 B_2)^{-1}\{U_{12} + N_1\ N_2 \delta_{12} P_0\}. \tag{42}$$

This equation displays the statistical fluctuation of $s$ as a sum of two parts, the atmospheric noise term $U_{12}$ reduced by negative feedback, and the photon-noise term $\delta_{12} P_0$ amplified by the feedback. To recapitulate the meaning of the symbols on the right of Eq. (42), $N$ is the feedback kernel, $B$ and $P_0$ are the optical response kernel and the unperturbed irradiance that appear in Eq. (9), $U_{12}$ is the atmospheric correlation function defined by Eq. (7), and $\delta_{12}$ is defined by Eq. (30).

## II. OPTIMIZATION OF THE FEEDBACK ALGORITHM

We take as the criterion for optimizing the system a weighted average of a quadratic function of path-length errors,

$$D = \sum\sum C(k_1, k_2)\,\langle\langle\, s(k_1, t)\, s(k_2, t)\,\rangle\rangle, \tag{43}$$

where $C(k_1, k_2)$ is an arbitrary symmetric positive-definite matrix. The value of $D$ is independent of the time $t$ at which the averages are taken. We define the system to be optimized when $D$ attains its minimum value. In the abbreviated notation of Sec. I, we have

$$D = \langle\langle sCs\rangle\rangle, \tag{44}$$

where $C$ now denotes the space–time kernel

$$C = C(k_1, k_2)\delta(t_1 - t_2). \tag{45}$$

Different choices for $C$ will emphasize different optical characteristics of the image.

As a standard of comparison by which to assess the significance of $D$, we take the quantity

$$D_0 = d^2(\mathrm{Tr}\,C), \tag{46}$$

where $d$ is one-tenth of an average optical wavelength. This $D_0$ is the value that $D$ would take if each path-length error $s(k)$ were an independent random variable with variance $d^2$. Because one-tenth of a wavelength is conventionally considered to be a tolerable error in optical systems, the condition

$$D \leq D_0 \tag{47}$$

is roughly what is required for the system to give diffraction-limited images of adequate quality.

The optical response of the system is sensitive only to path-length differences and not to absolute path lengths. A path-length error $s$ independent of $k$ produces no change of irradiance at the sensors. Formally, by Eq.

(9), if $u$ is the constant function

$$u(k, t) = 1, \tag{48}$$

then

$$Bu = 0. \tag{49}$$

Because constant path-length errors do not degrade the image quality, it is reasonable to make the criterion $D$ also insensitive to such errors. We therefore require the kernel $C$ to satisfy the condition

$$Cu = uC = 0. \tag{50}$$

When we substitute the final result Eq. (42) of Sec. I into Eq. (44), we obtain an expression for $D$ as the trace of a product of operators

$$D = \mathrm{Tr}\{C(1 - NB)^{-1}(U + NI_0N^t)(1 - B^tN^t)^{-1}\}, \tag{51}$$

where the trace indicates a summation over mirror-patch indices $k$ and detector indices $j$ and an integration over all time coordinates except one. The integration over the last time coordinate is omitted because the expression in Eq. (43) is independent of time. In Eq. (51), $B^t$ and $N^t$ are the operators adjoint to $B$ and $N$, defined by

$$B^t(y, x) = B(x, y), \quad N^t(x, y) = N(y, x), \tag{52}$$

and $I_0$ is the diagonal operator defined by

$$I_0(j_1, t_1, j_2, t_2) = \delta_{12}P_0(j_1), \tag{53}$$

with $\delta_{12}$ given by Eq. (30). Because absolute path lengths are unmeasurable and only path differences are significant, it is often convenient to define the atmospheric correlation function $U$ by

$$U(k_1, k_2, t_1 - t_2) = -\tfrac{1}{2}\langle\{a(k_1, t_1) - a(k_2, t_2)\}^2\rangle_a \tag{54}$$

rather than by Eq. (7). The expression in Eq. (54), unlike Eq. (7), can be directly related to observations of atmospheric turbulence. The two definitions of $U$ give identical results when substituted into Eq. (51), by virtue of Eqs. (49) and (50). It is important not to be misled by the minus sign that appears in Eq. (54). In spite of the minus sign, $U$ remains a positive-definite operator when it operates in the subspace of vectors orthogonal to $u$.

We now begin the process of minimization of $D$. Our preliminary objective is to prove the following duality theorem, which asserts that the feedback kernel $N(x, y)$ in an optimized system is proportional to the adjoint of the optical-response kernel $B(y, x)$.

*Theorem* 1. In an optimized system the feedback kernel $N$ has the form

$$N = KB^tI_0^{-1}, \tag{55}$$

where $K$ is a causal kernel still to be determined. The quantity $D$ defined by Eq. (43) has the value

$$D = \mathrm{Tr}\{C(1 - KT)^{-1}(U + KTK^t)(1 - TK^t)^{-1}\}, \tag{56}$$

with the symmetric kernel $T$ given by

$$T = B^tI_0^{-1}B. \tag{57}$$

*Remark* 1. We shall also prove that the kernel (55) automatically satisfies the condition (13) that we imposed in Sec. I to ensure that the unperturbed state of the system be an equilibrium state.

*Remark* 2. The main problem both in proving and applying Theorem 1 is to make sure that the inverse operator $I_0^{-1}$ in Eqs. (55) and (57) is correctly defined. The physical meaning of this $I_0^{-1}$ is clear; it means that we give greater importance in the feedback algorithm to photons detected in channels for which the average counting rate is smaller. The danger in this inverse weighting of photons is that unphysical singularities may arise in channels for which $I_0$ is zero. We therefore go through a careful discussion of definitions to make sure that singularities do not arise, either mathematically or operationally.

*Proof of Theorem* 1. The inverse operator $I_0^{-1}$ is well defined within the subspace $L_0$ of vectors $v(j)$ for which

$$v(j) = 0, \quad \text{when } P_0(j) = 0. \tag{58}$$

We define $L_1$ to be the space of vectors $v$ of the form

$$v = Ba \tag{59}$$

with $a$ arbitrary. In words, $L_1$ is the linear space of possible distributions of irradiance in the optical detectors produced in response to all possible atmospheric perturbations, according to the laws of classical electromagnetism. Statistical fluctuations of the photon counts will, in general, allow the real optical response to go outside the subspace $L_1$. It is essential to the argument that the optical-response kernel $B$ is a classically defined quantity that excludes photon-fluctuation effects from consideration.

We prove first that the space $L_0$ includes $L_1$. The classically computed irradiance $P(j)$, which appears in the definition (8) of $P_0$ and $B$, is a sum of squares of wave amplitudes

$$P(j) = \sum_p (A_p(j, s))^2, \tag{62}$$

that arise from different parts $p$ of the source when the path-length errors are described by the vector $s$. In Eq. (62), we make the assumption that different parts of the source radiate incoherently. The amplitudes $A_p$ are differentiable functions of the path-length errors $s(k)$, and Eq. (8) gives

$$P_0(j) = \sum_p [A_p(j, 0)]^2, \tag{63}$$

$$B_{jk} = \left(\frac{\partial}{\partial s(k)} P(j)\right)_{s=0}$$

$$= 2\sum_p A_p(j, 0)\left(\frac{\partial}{\partial s(k)} A_p(j, s)\right)_{s=0}. \tag{64}$$

From Eqs. (63) and (64), it follows that for each value of $j$, $P_0(j) = 0$ implies $B_{jk} = 0$ for all $k$. If, then, $v'$ is any vector orthogonal to $L_0$, we have

$$v'_j = 0, \quad \text{when } P_0(j) \neq 0; \tag{65}$$

therefore

$$v'_j = 0, \quad \text{when } B_{jk} \neq 0 \text{ for any } k, \tag{66}$$

and therefore $v'$ is orthogonal to all vectors in $L_1$ according to Eq. (59). This completes the proof that $L_0$ includes $L_1$.

The inverse $I_0^{-1}$ being defined in $L_1$ on all vectors of the form (59), the operator product

$$R = I_0^{-1} B \tag{67}$$

is defined everywhere and has an adjoint $R^t$. We then define the expression on the right-hand side of Eq. (55) by the convention

$$B^t I_0^{-1} = R^t, \tag{68}$$

which means that the operator $(B^t I_0^{-1})$ is by definition zero when it operates in the subspace for which $I_0^{-1}$ is undefined. The expression (57) for $T$ is also well defined, and

$$R^t I_0 = B^t, \quad R^t B = T, \tag{69}$$

are ordinary operator equations.

Let now $L_2$ be the linear space of vectors

$$v = Ta \tag{70}$$

with $a$ arbitrary. Within the space $L_2$, $T$ is a real symmetric matrix with strictly positive eigenvalues; therefore, $T$ has a partial inverse $A$, such that

$$TA = AT = Q \tag{71}$$

is the projection operator onto the space $L_2$. If $v'$ is any vector orthogonal to $L_2$, we have

$$Tv' = 0. \tag{72}$$

But $I_0^{-1}$ is well defined when it operates on the vector $Bv'$, and Eq. (72) implies

$$(Bv')^t I_0^{-1} (Bv') = (v')^t Tv' = 0, \tag{73}$$

which means that $Bv' = 0$ for all $v'$ orthogonal to $L_2$; that is,

$$B = BQ = BAT, \tag{74}$$

Given any feedback kernel $N$, we define the related causal kernel $K$ by

$$K = NBA. \tag{75}$$

To prove Theorem 1, we have to show that in an optimized system the relation (75) can be inverted so that $N$ is given by Eq. (55).

If we define $N'$ by

$$N = KR^t + N', \tag{76}$$

Theorem 1 states that $N' = 0$ in an optimized system. Equations (74) and (75) give

$$NB = NBAT = KT = KR^t B, \tag{77}$$

and therefore

$$N'B = 0. \tag{78}$$

Thus $N'$ is the part of $N$ that operates on vectors orthogonal to $L_1$. Equation (76) with Eqs. (69) and (78) implies

$$NI_0 N^t = KTK^t + N' I_0 N'^t. \tag{79}$$

Assembling Eqs. (51), (77), and (79), we find

$$D = \mathrm{Tr}\{C(1-KT)^{-1}(U + KTK^t + N' I_0 N'^t)(1-TK^t)^{-1}\}. \tag{80}$$

The $N'$ term in Eq. (76) contributes only to the photon noise in Eq. (80) and does not produce any useful feedback. If $N' \neq 0$, we can improve the performance of the system by setting $N' = 0$ and dropping the positive-definite term in $N'$ from Eq. (80). We thus obtain Eq. (55) and (56), and so Theorem 1 is proved.

*Proof of Remark* 1. We assume that the optical system is geometrically efficient, in the sense that all light entering the system is intercepted by optical detectors. This does not mean that the detectors are assumed to have an optical efficiency of 100%. The optical efficiency may have any value, provided that it is the same for all detectors. The assumption of geometrical efficiency means that atmospheric perturbations do not cause any light to be lost from the system, although they may transfer light flux from one detector to another. For any perturbation $s(k)$, the sum of irradiances $P(j)$ given by Eq. (8) should be equal to the sum of the unperturbed irradiances $P_0(j)$. This means that

$$B^t w = 0, \tag{81}$$

where $w$ is the vector of which all components are unity,

$$w(j) = 1. \tag{82}$$

The definition (53) of $I_0$ gives

$$I_0 w = P_0; \tag{83}$$

therefore Eq. (55) implies

$$NP_0 = KB^t w. \tag{84}$$

The condition (13) is then an immediate consequence of Eq. (81), and Remark 1 is proved.

Our next objective is to determine the time dependence of the feedback by optimizing the kernel $K$. We state the result of the optimization again as a formal theorem.

*Theorem* 2. In an optimized system with feedback kernel $N$ given by Eq. (55), $K$ is the unique causal kernel that satisfies the nonlinear integral equation

$$K + K^t - KTK^t + U = 0, \tag{85}$$

and the minimum value of the quantity $D$ given by Eqs. (43) or (56) is

$$D = -\mathrm{Tr}\{C(K + K^t)\}. \tag{86}$$

*Remark* 3. It is satisfactory that the optimum kernels $K$ and $N$ are independent of the arbitrary matrix $C$ that appears in the criterion of optimization $D$.

*Remark* 4. The causality condition to be satisfied by $K$ is

$$K = K(k_1, k_2, t_1 - t_2) = 0 \quad \text{for } t_1 < t_2. \tag{87}$$

When the trace on the right-hand side is written explicitly, Eq. (86) becomes

$$D = -\sum\sum C(k_2, k_1) K(k_1, k_2, 0+), \tag{88}$$

where $K(0+)$ means the limit of $K(t_1 - t_2)$ as $t_1 - t_2$ from above. Equation (88) shows a direct relation between the mean-square path-length errors of an optimized sys-

tem and the strength of the instantaneous negative feedback.

*Remark* 5. Theorem 2 contains three assertions: (a) existence of a causal $K$ that satisfies Eq. (85), (b) uniqueness of $K$, and (c) the statement that the $K$ so defined minimizes $D$. We shall prove these assertions in turn.

*Remark* 6. We already remarked upon the similarity between Eq. (85) and the Gelfand–Levitan equation.[8] Our proof of existence of $K$ is based on the analysis of the inverse scattering problem by Newton and Jost,[9] who generalized the Gelfand–Levitan theory to multichannel potentials.

*Proof of Existence.* If $W$ is any kernel, we define $P_a W$ to be the anticausal part of $W$. More precisely,

$$P_a W(t_1, t_2) = W(t_1, t_2) \quad \text{for } t_1 \leq t_2,$$
$$P_a W(t_1, t_2) = 0 \qquad \text{for } t_1 > t_2. \tag{89}$$

Note that $P_a$ operates on kernels and not on vectors. Now consider the linear integral equation

$$G^t + P_a U T G^t = P_a U \tag{90}$$

as an equation for the unknown anticausal kernel $G^t$. This is an equation of Fredholm type (see Riesz and Sz.-Nagy[10]). The operator $P_a UT$ is completely continuous when applied to $G^t$. Therefore the Fredholm alternative holds: either the inhomogeneous Eq. (90) has a solution, or the corresponding homogeneous equation

$$F + P_a U T F = 0 \tag{91}$$

has a nonzero solution $F$. But Eq. (91) implies for an anticausal $F$

$$\mathrm{Tr}[F^t T F] + \mathrm{Tr}[F^t T U T F] = 0 \tag{92}$$

and, since both $T$ and $U$ are positive definite, Eq. (92) implies $TF = 0$. Then Eq. (91) gives $F = 0$, so we have proved the existence of $G^t$, satisfying Eq. (90). We define the kernel $K$ by

$$K = G^t + U T G^t - U, \tag{93}$$

so that Eq. (90) becomes

$$P_a K = 0. \tag{94}$$

Thus $K$ is a causal kernel; we now have only to prove that it satisfies Eq. (85).

The equation adjoint to (93) is

$$K^t = G + G T U - U. \tag{95}$$

We multiply Eq. (93) on the left-hand side by $(1 - GT)$, multiply Eq. (95) on the right-hand side by $(1 - TG^t)$, and subtract. The result is

$$G + K - GTK = G^t + K^t - K^t T G^t. \tag{96}$$

But the left-hand side of Eq. (96) is causal, whereas the right-hand side is anticausal. Therefore both sides are zero, and we have

$$(1 - GT)(1 - KT) = (1 - TG)(1 - TK) = 1, \tag{97}$$

$$G = -(1 - KT)^{-1} K, \quad K = -(1 - GT)^{-1} G, \tag{98}$$

the inverses here being well defined by virtue of Eq. (97). When we substitute $G$ from Eq. (98) into (95), the result is Eq. (85). This completes the proof of the existence part of Theorem 2.

*Proof of Uniqueness.* Since $T$ and $U$ are positive definite, the operator $(1 + UT)$ has an inverse. Now if $K$ is any solution of Eq. (85), we have

$$(1 - KT)(1 - K^t T) = 1 + UT, \tag{99}$$

and so the inverse $(1 - KT)^{-1}$ also exists. If $K_1$ and $K_2$ are any two solutions of Eq. (85), we find by equating the two expressions for $U$,

$$(1 - K_1 T)^{-1}(K_1 - K_2) = (K_2^t - K_1^t)(1 - TK_2^t)^{-1}. \tag{100}$$

If $K_1$ and $K_2$ are both causal, the left-hand side of Eq. (100) is causal and the right-hand side is anticausal, so that both sides must be zero. This proves the uniqueness of $K$.

*Proof that D is a minimum.* Let $K$ be the solution of Eq. (85) and let $K'$ be any other causal kernel for which the value $D'$ of $D$ is defined by Eq. (56). We define a causal kernel $H$ by

$$H = (1 - K'T)^{-1}(K' - K), \tag{101}$$

so that

$$(1 - K'T)^{-1} = (1 + HT)(1 - GT), \tag{102}$$

$$(1 - K'T)^{-1} K' = H - (1 + HT)G. \tag{103}$$

Using Eq. (102) and (103), we eliminate $K'$ from Eq. (56), and express $D'$ in terms of $G$ and $H$. We eliminate $U$ from $D'$ by using the equation

$$(1 - GT)U(1 - TG^t) = G + G^t - GTG^t, \tag{104}$$

which is obtained by substituting for $K$ from Eq. (98) into Eq. (93). The resulting formula for $D'$ is

$$D' = \mathrm{Tr}\{C(G + G^t + HTG + G^t T H^t + HTH^t)\}. \tag{105}$$

The purpose of the substitution Eq. (101) was to reduce $D'$ to a quadratic form in the unknown $H$. Now

$$\mathrm{Tr}\{CHTG\} = \mathrm{Tr}\{CG^t T H^t\} = 0, \tag{106}$$

because both $G$ and $H$ are causal kernels while $C$ and $T$ are instantaneous. Therefore $D'$ depends on $H$ only through the quadratic term which is positive definite. So we have proved that $D'$ attains its minimum value when $H = 0$ and $K' = K$. The value of $D'$ at the minimum is

$$D = \mathrm{Tr}\{C(G + G^t\}. \tag{107}$$

We have also

$$\mathrm{Tr}\{CGTK\} = \mathrm{Tr}\{CK^t T G^t\} = 0, \tag{108}$$

because $G$ and $K$ are both causal. Because both sides of Eq. (96) are zero, the value of $D$ given by Eq. (107) coincides with Eq. (86). This completes the proof of Theorem 2.

## III. APPLICATION TO PARTICULAR SYSTEMS

The theory in Secs. I and II can be applied to active optical systems of many kinds. We have carried the analysis through in detail for three special cases, (1) a simple stellar interferometer of the type developed by

Currie,[11] (2) a conventional telescope with detectors in the image plane, as described by Muller and Buffington[2] and by Bridges *et al.*,[3] and (3) a system that uses an array of interferometers, as described by Hardy *et al.*[4] and by Dicke.[12] We take the criterion of successful operation of the system to be Eq. (47), with $D$ and $D_0$ defined by Eqs. (43) and (46). According to this criterion, a successful system should give angular resolution equal to the diffraction limit. In each case, we suppose the feedback algorithm optimized according to the prescriptions of Sec. II. The following is a brief summary of the results.

(1) For the simple interferometer, we have as the condition for successful operation

$$\phi p^2 \eta^2 \tau e > a^2 , \tag{109}$$

where $\phi$ is the incoming photon flux (photon number per unit area per unit time), $p$ is the diameter of each of the two receiver apertures (assumed small enough so that atmospheric phase differences within each patch are unimportant), $\eta$ is the fraction of the light contributing to optical structure at the desired resolution (equal to the fractional modulation of the observed interference pattern), $\tau$ is the average autocorrelation time and $a$ the root-mean-square amplitude, in radians, of the atmospheric phase difference between the two interfering beams, and $e$ is the fractional efficiency of the detectors.

(2) For the system with image-plane detection, the criterion for success is

$$\phi p^2 \eta^2 \tau e > (A/p^2) , \tag{110}$$

with $\phi$, $\eta$, $e$ defined as before, whereas $p$ is now the seeing-patch diameter (diameter of the patches on the mirror that are independently controlled), $\tau$ is the autocorrelation time of the phase differences between neighboring patches, and $A$ is the total area of the mirror.

(3) For a system with detection by multiple interferometers that monitor phase differences over all of the receiving area, the criterion for success is

$$\phi p^2 \eta^2 \tau e > 1, \tag{111}$$

with the symbols defined as in case (2). In each of Eqs. (109), (110), and (111), a numerical coefficient of the order of unity has been dropped, the precise value of the coefficient depending on details of the system design and of the atmospheric fluctuation spectrum. The quantity that appears on the left-hand side in each case is the effective number of photons that are detected per resolution element of the system in space and time.

The criterion (110) for image-plane detection has the aperture area on the right-hand side. This means that image-plane detection is undesirable for large systems. The inefficiency of image-plane detection arises from the mixing of photons from all parts of the mirror at each of the detectors. When photons are mixed, information is lost that cannot be recovered in subsequent data processing. The condition (111) shows that with interferometric detection the limiting photon flux is independent of the size of the system. If an object is bright enough to fulfill condition (111), it can be observed with angular resolution limited only by the size of the instruments that it is feasible to build. It is likely that, in optical astronomy as in radio astronomy, the very-high-resolution instruments of the future will be interferometric arrays rather than single dishes.

If we substitute into Eq. (111) numerical values appropriate to conditions of good seeing at ground-based observatories, namely

$$p = 10 \text{ cm}, \quad \tau = 10^{-2} \text{ s} , \tag{112}$$

with a source possessing sharply defined optical structure ($\eta = 1$), and assume a detection efficiency $e = 0.1$, we find the limiting source brightness to be

$$\phi = 10 \text{ cm}^{-2} \text{ s}^{-1} . \tag{113}$$

This means that objects brighter than magnitude 14 should be observable with arbitrarily high resolution, under conditions of good seeing. The numerical estimate is of course very rough, and is intended only as a guide to the planning of experimental systems, rather than as a prediction of their performance.

ACKNOWLEDGMENTS

I am grateful to Dr. Richard Muller for many useful discussions, and to Dr. Hal Lewis for originally suggesting the problem. This research was partially supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0370 while the author was a consultant to Stanford Research Institute.

*On leave of absence from the Institute for Advanced Study, Princeton, N. J. 08540.

[1]H. W. Babcock, Publ. Astron. Soc. Pac. 65, 229 (1953).
[2]R. A. Muller and A. Buffington, J. Opt. Soc. Am. 64, 1200 (1974).
[3]W. B. Bridges, P. T. Brunner, S. P. Lazzara, T. A. Nussmeier, T. R. O'Meara, J. A. Sanguinet, and W. P. Brown, Jr., Appl. Opt. 13, 291 (1974).
[4]J. W. Hardy, J. Feinleib, and J. C. Wyant, in *Digest of Technical Papers, Topical Meeting on Optical Propagation through Turbulence*, sponsored by the OSA, Boulder, Colo., July 1974, paper ThB1.
[5]L. Miller, W. P. Brown, Jr., J. A. Jenney, and T. R. O'Meara, Ref. 4, paper ThB2.
[6]R. F. Ogrodnik, Ref. 4, paper ThB4.
[7]J. E. Pearson, W. B. Bridges, L. S. Horwitz, T. J. Walsh, and R. F. Ogrodnik, Ref. 4, paper ThB5.
[8]I. M. Gelfand and B. M. Levitan, Izv. Akad. Nauk SSSR, Ser. Mat. 15, 309 (1951).
[9]R. G. Newton and R. Jost, Nuovo Cimento 1, 590 (1955).
[10]F. Riesz and B. Sz.-Nagy, *Functional Analysis*, translated by L. F. Boron (Ungar, New York, 1955), p. 161.
[11]D. G. Currie, Ref. 4, paper WB9.
[12]R. H. Dicke (private communication).

# APPENDIX F

Reprinted from *Journal of the Optical Society of America*, Vol. 67(3), pp. 370-375 (March 1977).

# Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements

David L. Fried
*Optical Science Consultants, P. O. Box 446, Placentia, California 92670*
(Received 17 July 1976)

The problem of fitting a wave-front distortion estimate to a (single-instant) set of phase-difference measurements has been formulated as an unweighted least-square problem. The least-square equations have been developed as a set of simultaneous equations for a square array of phase-difference sensors, with phase estimates at the corner of each measurement element. (This corresponds to the standard Hartmann configuration and to one version of a shearing interferometer of a predetection compensation wave-front sensor.) The noise dependence in the solution of the simultaneous equations is found to be expressible in terms of the solution to a particular version of the measurement inputs to the simultaneous equation, a sort of "Green's-function" solution. The noise version of the simultaneous equations is solved using relaxation techniques for array sizes from $4 \times 4$ to $40 \times 40$ phase estimation points, and the mean-square wave-front error calculated as a function of the mean-square phase-difference measurement error. It is found that the results can be approximated within a fraction of a percent accuracy by $\langle(\delta\Phi)^2\rangle = 0.6558[1 + 0.2444 \ln(N^2)]\sigma_{pd}^2$, where $\langle(\delta\Phi)^2\rangle$ is the mean-square error (rad$^2$) in the estimation of the wave-front distortion, for a square array consisting of $N^2$ square subaperture elements over which two phase-difference measurements are made—one phase difference across the $x$ dimension and the other difference across the $y$ dimension. Here $\sigma_{pd}^2$ is the mean-square error (rad$^2$) in each phase-difference measurement.

## INTRODUCTION

In this paper we shall address the question of the performance of a processor which, given a set of phase-difference measurements across a large aperture, has to formulate a corresponding estimate of the distorted wave front over that aperture. This processor is the key to certain types of adaptive optics systems, particularly those that utilize a pseudo-Hartmann or a shearing-interferometer-type sensor.[1-8]

We measure the performance of the processor in terms of the rms error in the estimated wave front, $\langle(\delta\Phi)^2\rangle$. We assume that the mean-square error in the input phase-difference measurements, $\sigma_{pd}^2$, is known and will see that it controls the error in the processor output. We further assume that since there are more phase-difference measurement inputs than points at which the processor is to estimate the phase difference (nearly twice as many), a least-square error estimation procedure will be used.

## DEFINITION OF ELEMENTS ON THE APERTURE PLANE

For the purposes of our analysis, we shall consider an aperture in the form of a square array of square regions, arranged like a checkerboard. We shall restrict our attention to arrays consisting of an odd number of elements, i.e., the array size will be

$$N^2 = (2n+1)^2 . \tag{1}$$

We consider the element to be a region across which a phase difference is estimated. For the pseudo-Hartmann measurement technique, the measured phase difference between opposite sides of the element would correspond to the measured angle-of-arrival times the element dimension, converted from length into radians of phase. We would write

$$\phi_x = kd\theta_x , \tag{2a}$$

$$\phi_y = kd\theta_y , \tag{2b}$$

where $k = 2\pi/\lambda$ is the optical wave number, $d$ is the length of a side of the element, $\theta_x$ and $\theta_y$ are the measured angle-of-arrival components, and $\phi_x$ and $\phi_y$ are the phase differences between the two sides of the square element. If a shearing interferometer is used to measure the phase difference, then it is convenient to consider the shear distance to be $\delta$. Then

$$\phi_x = \Delta\phi_x(d/\delta) , \tag{3}$$

$$\phi_y = \Delta\phi_y(d/\delta) , \tag{4}$$

where $\Delta\phi_x$ and $\Delta\phi_y$ are the measured phase differences for shear of distance $\delta$ in the $x$ and $y$ directions, respectively.

It is not our intention to take up here the question of the relative merits of each of these phase-difference measurement options. There are considerations concerned with photon shot noise effects and the fact that the measurements are only approximately the phase difference across the aperture element rather than the true phase difference between the two sides as a result of spatial quantization in our measurements. At this point we simply wish to know the rms error in the estimated instantaneous wave-front distortion, assuming that we have meaningful measurements of the phase difference, $\phi_x$ and $\phi_y$, between the two sides of the aperture element.

We shall define the instantaneous wave-front distortion by phase estimates at each of the four corners of each aperture element. It is convenient to denote the element by the notation $(i,j)$ with $-n \leq i \leq n$ and $-n \leq j \leq n$. We shall let $\phi_{x,i,j}$ and $\phi_{y,i,j}$ denote the phase difference between the two sides of the $(i,j)$th aperture element. We shall use the notation $\Phi_{p,q}$ to denote the estimated phase (matching the instantaneous wave-front distortion) at the corner common to the $(i = p+\frac{1}{2}, j = q+\frac{1}{2})$, $(i = p+\frac{1}{2}, j = q-\frac{1}{2})$, $(i = p-\frac{1}{2}, j = q+\frac{1}{2})$, and $(i = p-\frac{1}{2}, j = q-\frac{1}{2})$ aperture elements. We let $p$ and $q$ run through the values $p = -n-\frac{1}{2}, -n+\frac{1}{2}, -n+\frac{3}{2}, \ldots, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots, n-\frac{3}{2}, n-\frac{1}{2}, n+\frac{1}{2}$, and similarly,

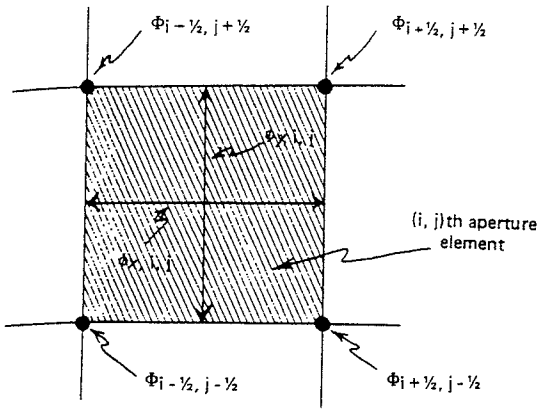



FIG. 1. Representation of the orientation and location of phase differences and phase estimates for the $(i,j)$th aperture element.

$q = -n-\frac{1}{2}, -n+\frac{1}{2}, -n+\frac{3}{2}, \ldots, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots, n-\frac{3}{2}, n-\frac{1}{2}, n+\frac{1}{2}$. The relationship between $\phi_{x,i,j}$, $\phi_{y,i,j}$, and the various $\Phi_{p,q}$ for the $(i,j)$th aperture element is indicated in Fig. 1.

With the quantities of concern to us in our problem thus defined, we now turn our attention to the formulation of a method of calculating the $\Phi_{p,q}$ terms that make up the instantaneous wave-front distortion estimate, given the instantaneous measurements of the phase differences, $\phi_{x,i,j}$ and $\phi_{y,i,j}$. This is treated in the next section. After that we shall turn our attention to matters related to the estimation of the rms error to be expected in $\Phi_{p,q}$, given the rms error in $\phi_{x,i,j}$ and in $\phi_{y,i,j}$.

## LEAST-SQUARE ESTIMATION OF $\Phi_{p,q}$

We note that there are two phase-difference measurements, $\phi_{x,i,j}$ and $\phi_{y,i,j}$, for the $(i,j)$th aperture element, so that for $N^2$ aperture elements there will be $2N^2$ measurements in each instant. We also note that for $N^2$ aperture elements there are $(N+1)^2$ values of $\Phi_{p,q}$ to be estimated. The $\Phi_{p,q}$ are degrees-of-freedom to be adjusted to match the constraints provided by the $\phi_{x,i,j}$ and $\phi_{y,i,j}$ measurements. We see that for $N \geq 3$ there are more constraints than adjustable degrees-of-freedom, and so a least-square fit to the data is called for.

The phase differences across the $(i,j)$th aperture element are related to the phase estimates by the equations

$$\bar{\phi}_{x,i,j} = \tfrac{1}{2}(\Phi_{i+(1/2),j+(1/2)} + \Phi_{i+(1/2),j-(1/2)}) - \tfrac{1}{2}(\Phi_{i-(1/2),j+(1/2)} + \Phi_{i-(1/2),j-(1/2)}), \quad (5a)$$

$$\bar{\phi}_{y,i,j} = \tfrac{1}{2}(\Phi_{i+(1/2),j+(1/2)} + \Phi_{i-(1/2),j+(1/2)}) - \tfrac{1}{2}(\Phi_{i+(1/2),j-(1/2)} + \Phi_{i-(1/2),j-(1/2)}), \quad (5b)$$

where $\bar{\phi}_{x,i,j}$ and $\bar{\phi}_{y,i,j}$ are defined by Eqs. (5a) and (5b), with the $\Phi_{p,q}$ chosen so as to minimize the discrepancy between the $\bar{\phi}_{x,i,j}$ and $\bar{\phi}_{y,i,j}$ as compared to the $\phi_{x,i,j}$ and $\phi_{y,i,j}$. This discrepancy can be defined in unweighted least-square error terms as

$$\Delta = \sum_{i,j} [(\phi_{x,i,j} - \bar{\phi}_{x,i,j})^2 + (\phi_{y,i,j} - \bar{\phi}_{y,i,j})^2]. \quad (6)$$

We wish to choose the $\Phi_{p,q}$ so as to minimize the squared error, $\Delta$.

It should be noted that if we were only interested in short-exposure imagery, for which the average tilt across the aperture has no effect, then Eq. (6) would have to be modified to reflect this. This would be accomplished by subtracting the average over $i', j'$ of $\phi_{x,i',j'} - \bar{\phi}_{x,i',j'}$ from each $(\phi_{x,i,j} - \bar{\phi}_{x,i,j})$ term before squaring, and similarly subtracting from the $(\phi_{y,i,j} - \bar{\phi}_{y,i,j})$ terms before squaring. Since, however, we are interested in long-exposure resolution, failure to consider an average tilt error would be erroneous. We therefore have left the average tilt error in our expression for $\Delta$ as formulated in Eq. (6).

Our selection of the estimated phase values is based on the requirement that $\Delta$ be minimized. Thus we write

$$\frac{\partial \Delta}{\partial \Phi_{p,q}} = 0 \quad (7)$$

for all values of $(p,q)$, thus establishing a set of $(N+1)^2$ simultaneous equations to be solved for the $(N+1)^2$ unknown values, $\Phi_{p,q}$. If we substitute Eq. (6) into Eq. (7), we get for all $(p,q)$

$$2\sum_{i,j}\left((\phi_{x,i,j} - \bar{\phi}_{x,i,j})\frac{\partial \bar{\phi}_{x,i,j}}{\partial \Phi_{p,q}} + (\phi_{y,i,j} - \bar{\phi}_{y,i,j})\frac{\partial \bar{\phi}_{y,i,j}}{\partial \Phi_{p,q}}\right) = 0. \quad (8)$$

Making use of Eqs. (5a) and (5b), we see that

$$2\frac{\partial \bar{\phi}_{x,i,j}}{\partial \Phi_{p,q}} = \delta(p, i+\tfrac{1}{2})\delta(q, j+\tfrac{1}{2}) + \delta(p, i+\tfrac{1}{2})\delta(q, j-\tfrac{1}{2}) - \delta(p, i-\tfrac{1}{2})\delta(q, j+\tfrac{1}{2}) - \delta(p, i-\tfrac{1}{2})\delta(q, j-\tfrac{1}{2}), \quad (9a)$$

$$2\frac{\partial \bar{\phi}_{y,i,j}}{\partial \Phi_{p,q}} = \delta(p, i+\tfrac{1}{2})\delta(q, j+\tfrac{1}{2}) + \delta(p, i-\tfrac{1}{2})\delta(q, j+\tfrac{1}{2}) - \delta(p, i+\tfrac{1}{2})\delta(q, j-\tfrac{1}{2}) - \delta(p, i-\tfrac{1}{2})\delta(q, j-\tfrac{1}{2}). \quad (9b)$$

The delta functions here are intended to denote Kronecker deltas, according to which

$$\delta(a,b) = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b. \end{cases} \quad (10)$$

If we substitute Eqs. (9a) and (9b) into Eq. (8), and exploit the property of the Kronecker deltas to allow us to carry out the $(i,j)$ summation, we obtain for all $(p,q)$

$$(\phi_{x,p-(1/2),q-(1/2)} - \bar{\phi}_{x,p-(1/2),q-(1/2)}) + (\phi_{x,p-(1/2),q+(1/2)} - \bar{\phi}_{x,p-(1/2),q+(1/2)}) - (\phi_{x,p+(1/2),q-(1/2)} - \bar{\phi}_{x,p+(1/2),q-(1/2)}) - (\phi_{x,p+(1/2),q+(1/2)} - \bar{\phi}_{x,p+(1/2),q+(1/2)}) + (\phi_{y,p-(1/2),q-(1/2)} - \bar{\phi}_{y,p-(1/2),q-(1/2)}) + (\phi_{y,p+(1/2),q-(1/2)} - \bar{\phi}_{y,p+(1/2),q-(1/2)}) - (\phi_{y,p-(1/2),q+(1/2)} - \bar{\phi}_{y,p-(1/2),q+(1/2)}) - (\phi_{y,p+(1/2),q+(1/2)} - \bar{\phi}_{y,p+(1/2),q+(1/2)}) = 0. \quad (11)$$

If we substitute Eqs. (5a) and (5b) into Eq. (11) and collect the various $\Phi$ terms, we get for all $(p, q)$ for which $p \neq \pm(n+\frac{1}{2})$ and $q \neq \pm(n+\frac{1}{2})$

$$4\Phi_{p,q} - \Phi_{p-1,q-1} - \Phi_{p-1,q+1} - \Phi_{p+1,q-1} - \Phi_{p+1,q+1} = \phi_{x,p-(1/2),q-(1/2)} + \phi_{x,p-(1/2),q+(1/2)} - \phi_{x,p+(1/2),q-(1/2)} - \phi_{x,p+(1/2),q+(1/2)}$$
$$+ \phi_{y,p-(1/2),q-(1/2)} + \phi_{y,p+(1/2),q-(1/2)} - \phi_{y,p-(1/2),q+(1/2)} - \phi_{y,p+(1/2),q+(1/2)} . \quad (12)$$

For the various cases in which $p$ and/or $q$ equal $\pm(n+\frac{1}{2})$, certain terms in Eq. (12) would be missing since they pertain to aperture elements that are not part of the $(2n+1)^2$ elements making up the total array. There are eight variant forms of Eq. (12) governing the cases for which $p$ and/or $q$ equal $\pm(n+\frac{1}{2})$. These are most compactly represented if we define the functions $f(i,j)$ and $g(i,j)$ as

$$f(i,j) = \begin{cases} 1, & \text{if } |i| \leq n+\frac{1}{2} \text{ and } |j| \leq n+\frac{1}{2}, \\ 0, & \text{otherwise;} \end{cases} \quad (13)$$

$$g(i,j) = \begin{cases} 4, & \text{if } |i| < n+\frac{1}{2} \text{ and } |j| < n+\frac{1}{2}, \\ 1, & \text{if } |i| = n+\frac{1}{2} \text{ and } |j| = n+\frac{1}{2}, \\ 2, & \text{otherwise.} \end{cases} \quad (14)$$

Making use of these functions, we can write the general version of Eq. (12), which is valid for *all* values of $(p, q)$ as

$$g(p,q)\Phi_{p,q} - f(p-1,q-1)\Phi_{p-1,q-1} - f(p-1,q+1)\Phi_{p-1,q+1} - f(p+1,q-1)\Phi_{p+1,q-1} - f(p+1,q+1)\Phi_{p+1,q+1}$$
$$= f(p-\tfrac{1}{2},q-\tfrac{1}{2})(\phi_{x,p-(1/2),q-(1/2)} + \phi_{y,p-(1/2),q-(1/2)}) + f(p-\tfrac{1}{2},q+\tfrac{1}{2})(\phi_{x,p-(1/2),q+(1/2)} - \phi_{y,p-(1/2),q+(1/2)})$$
$$+ f(p+\tfrac{1}{2},q-\tfrac{1}{2})(-\phi_{x,p+(1/2),q-(1/2)} + \phi_{y,p+(1/2),q-(1/2)}) + f(p+\tfrac{1}{2},q+\tfrac{1}{2})(-\phi_{x,p+(1/2),q+(1/2)} - \phi_{y,p+(1/2),q+(1/2)}) . \quad (15)$$

Equation (15) represents $(2n+2)^2$ simultaneous equations giving the values of the $(2n+2)^2$ values of $\Phi_{p,q}$, based on the $2(2n+1)^2$ values of the measured phase differences $\phi_{x,i,j}$ and $\phi_{y,i,j}$. These simultaneous equations represent a least-square fit of the instantaneous estimated wave-front distortion to the phase-difference measurements.

With the least-square solution formulated in terms of Eq. (15), we are now ready to turn our attention to the evaluation of the mean-square instantaneous wave-front distortion error to be expected as a consequence of some rms phase-difference measurement error. We take up the formulation of this problem in the next section.

## rms WAVE-FRONT ERROR—FORMULATION

We start our evaluation by noting that Eq. (15) represents a set of $(2n+2)^2$ linear simultaneous equations. The fact that the equations are linear means that if $\{(\Phi_{p,q})_1\}$ are a set of $(2n+2)^2$ solutions associated with the set $\{(\phi_{x,i,j})_1; (\phi_{y,i,j})_1\}$ of $2(2n+1)^2$ measurements, and if $\{(\Phi_{p,q})_2\}$ is the set of solutions corresponding to the measurement set $\{(\phi_{x,i,j})_2; (\phi_{y,i,j})_2\}$, then the set of solutions $\{(\Phi_{p,q})_1 + (\Phi_{p,q})_2\}$ corresponds to the measurement set $\{(\phi_{x,i,j})_1 + (\phi_{x,i,j})_2; (\phi_{y,i,j})_1 + (\phi_{y,i,j})_2\}$. Recognizing the implication of this linearity, we perform the following decomposition of the measurements and calculated instantaneous wave-front distortion:

$$\phi_{x,i,j} = \bar{\phi}_{x,i,j} + \delta\phi_{x,i,j}, \quad (16a)$$

$$\phi_{y,i,j} = \bar{\phi}_{y,i,j} + \delta\phi_{y,i,j}, \quad (16b)$$

where $\bar{\phi}_{x,i,j}$ and $\bar{\phi}_{y,i,j}$ are the correct values that the measurements should yield (if there were no noise), and $\delta\phi_{x,i,j}$ and $\delta\phi_{y,i,j}$ are the noise contributions to the measurements. We write the calculated instantaneous wave-front distortion

$$\Phi_{p,q} = \bar{\Phi}_{p,q} + \delta\Phi_{p,q}, \quad (17)$$

where $\bar{\Phi}_{p,q}$ are the correct values of the instantaneous wave-front distortion, which would be calculated by using the $\bar{\phi}_{x,i,j}$ and $\bar{\phi}_{y,i,j}$, so that $\delta\Phi_{p,q}$ is the error in the calculated instantaneous wave-front distortion due to the measurement noise. It follows from the linearity of the simultaneous equations that $\{\delta\Phi_{p,q}\}$ is the solution of Eq. (15) if $\{\delta\phi_{x,i,j}; \delta\phi_{y,i,j}\}$ are taken as the measurements to be input into Eq. (15).

The implication of the linearity of the simultaneous equations can be pushed even further. If we take all $\{(\Phi_{p,q})_{x,i,j}\}$ to be the solution to Eq. (15) when the $2(2n+1)^2$ measurements are replaced by zero, except for the $\phi_{x,i,j}$ measurement which is replaced by unity, and if $\{(\Phi_{p,q})_{y,i,j}\}$ is the solution when all the measurements are zero except the $\phi_{y,i,j}$ measurement, then it follows from the linearity of the equations that

$$\delta\Phi_{p,q} = \sum_{i,j} \delta\phi_{x,i,j}(\Phi_{p,q})_{x,i,j} + \delta\phi_{y,i,j}(\Phi_{p,q})_{y,i,j} . \quad (18)$$

The mean-square wave-front distortion can be written

$$\langle(\delta\Phi)^2\rangle = (2n+1)^{-2} \sum_{p,q} \tfrac{1}{4} g(p,q) \langle(\delta\Phi_{p,q})^2\rangle , \quad (19)$$

where we have weighted the $(p, q)$ locations on the interior of the array covering the aperture at unity, those at the four corners at one-quarter, and those on the edge but not at the corner at one-half. With this weighting to appropriately discount $(p, q)$ boundary points, the factor $(2n+1)^{-2}$ provides the correct normalization for the $(2n+2)^2$ term summation.

The evaluation of $\langle(\delta\Phi_{p,q})^2\rangle$ follows directly from Eq. (18). We shall make use of the statistical independence of the measurement noise values, and the fact that each measurement has the same rms phase-difference error, $\sigma_{pd}$. We write this in the form

$$\langle\delta\phi_{x,i,j}\delta\phi_{x,i',j'}\rangle = \sigma_{pd}^2 \delta(i,i')\delta(j,j'), \quad (20a)$$

$$\langle\delta\phi_{y,i,j}\delta\phi_{y,i',j'}\rangle = \sigma_{pd}^2 \delta(i,i')\delta(j,j'), \quad (20b)$$

$$\langle \delta\phi_{x,i,j}\delta\phi_{y,i',j'}\rangle = 0\,, \tag{20c}$$

where $\delta(a, b)$ is the Kronecker delta, defined in Eq. (10). If we substitute Eq. (18) in the evaluation of $\langle(\delta\Phi_{p,q})^2\rangle$, we can cast the result in the form of a quadruple sum on $i, i', j, j'$, but by making use of Eqs. (20a), (20b), and (20c) this can be reduced to a double sum on $i, j$. Thus we obtain

$$\langle(\delta\Phi_{p,q})^2\rangle = \sigma_{pd}^2 \sum_{i,j}\left\{[(\Phi_{p,q})_{x,i,j}]^2 + [(\Phi_{p,q})_{y,i,j}]^2\right\}. \tag{21}$$

Now if we substitute Eq. (21) into Eq. (19) and interchange the order of the $(p, q)$ summation and the $(i, j)$ summation, we get for the mean-square instantaneous wave-front distortion error

$$\langle(\delta\Phi)^2\rangle = (2n+1)^{-2}\sum_{i,j}\left(\sigma_{pd}^2 \sum_{p,q} \tfrac{1}{4} g(p, q)\{[(\Phi_{p,q})_{x,i,j}]^2 + [(\Phi_{p,q})_{y,i,j}]^2\}\right). \tag{22}$$

At this point as a matter of practical simplification, we introduce the assumption that the sum over $(p, q)$ has a value which is very nearly independent of $(i, j)$, i.e., we assume that the wave-front distortion mean-square value over the aperture due to a single measurement error is independent of where the error is located, and also of whether it is a $\phi_{x,i,j}$ error or a $\phi_{y,i,j}$ error. This allows us to replace $[(\Phi_{p,q})_{x,i,j}]^2$ and $[(\Phi_{p,q})_{y,i,j}]^2$ in Eq. (22) with the corresponding values for $i = 0$, $j = 0$, with no need to distinguish between subscript $x$ and subscript $y$. We write

$$(\Phi_{p,q})_{00} = (\Phi_{p,q})_{x,i,j} \quad \text{for } i = 0,\ j = 0\,, \tag{23}$$

and now can cast Eq. (22) in the form

$$\langle(\delta\Phi)^2\rangle = (2n+1)^{-2}\sum_{i,j}\left(2\sigma_{pd}^2 \sum_{p,q} \tfrac{1}{4} g(p, q)[(\Phi_{p,q})_{00}]^2\right). \tag{24}$$

We note that the quantity in the large parentheses is independent of $i, j$, so that when the $(i, j)$ sum is carried out, it simply gives a factor of $(2n+1)^2$ (since $i$ and $j$ each run over the $2n+1$ values from $-n$ to $+n$). Thus Eq. (24) is reduced to the form

$$\langle(\delta\Phi)^2\rangle = 2\sigma_{pd}^2\,\mathfrak{N}, \tag{25}$$

where

$$\mathfrak{N} = \sum_{p,q} \tfrac{1}{4} g(p, q)[(\Phi_{p,q})_{00}]^2\,. \tag{26}$$

The problem of evaluation of the mean-square instantaneous wave-front distortion error $\langle(\delta\Phi)^2\rangle$ associated with an rms phase difference measurement error $\sigma_{pd}$ is seen to reduce to the evaluation of $\mathfrak{N}$. Results of this evaluation for various size aperture arrays are presented in the next section.

## rms WAVE-FRONT ERROR—EVALUATION

The evaluation of the rms instantaneous wave-front distortion error due to phase-difference measurement errors requires the evaluation of $\mathfrak{N}$, and as we see from Eq. (26), this is basically the problem of evaluating the set $\{(\Phi_{p,q})_{00}\}$. The evaluation of the $(2n+2)^2$ terms in the set $\{(\Phi_{p,q})_{00}\}$ requires the solution of Eq. (15) with all the $\phi_{x,i,j}$ and $\phi_{y,i,j}$ replaced by zero, except $\phi_{x,0,0}$, which we replace by unity. For the four cases defined by $p = \pm\frac{1}{2}$, $q = \pm\frac{1}{2}$, the right-hand side of Eq. (15) does not vanish. Otherwise it does. Thus Eq. (15) reduces to the five specific cases of

$$(\Phi_{1/2,1/2})_{00} = [g(\tfrac{1}{2}, \tfrac{1}{2})]^{-1}\{f(-\tfrac{1}{2}, \tfrac{1}{2})(\Phi_{-1/2,-1/2})_{00} + f(-\tfrac{1}{2}, \tfrac{3}{2})(\Phi_{-1/2,3/2})_{00} + f(\tfrac{3}{2}, -\tfrac{1}{2})(\Phi_{3/2,-1/2})_{00} + f(\tfrac{3}{2}, \tfrac{3}{2})(\Phi_{3/2,3/2})_{00} + 1\}, \tag{27a}$$

$$(\Phi_{1/2,-1/2})_{00} = [g(\tfrac{1}{2}, -\tfrac{1}{2})]^{-1}\{f(-\tfrac{1}{2}, -\tfrac{3}{2})(\Phi_{-1/2,-3/2})_{00} + f(-\tfrac{1}{2}, \tfrac{1}{2})(\Phi_{-1/2,1/2})_{00} + f(\tfrac{3}{2}, -\tfrac{3}{2})(\Phi_{3/2,-3/2})_{00} + (\tfrac{3}{2}, \tfrac{1}{2})(\Phi_{3/2,1/2})_{00} + 1\}, \tag{27b}$$

$$(\Phi_{-1/2,1/2})_{00} = [g(-\tfrac{1}{2}, \tfrac{1}{2})]^{-1}\{f(-\tfrac{3}{2}, -\tfrac{1}{2})(\Phi_{-3/2,-1/2})_{00} + f(-\tfrac{3}{2}, \tfrac{3}{2})(\Phi_{-3/2,3/2})_{00} + f(\tfrac{1}{2}, -\tfrac{1}{2})(\Phi_{1/2,-1/2})_{00} + f(\tfrac{1}{2}, \tfrac{3}{2})(\Phi_{1/2,3/2})_{00} - 1\}, \tag{27c}$$

$$(\Phi_{-1/2,-1/2})_{00} = [g(-\tfrac{1}{2}, -\tfrac{1}{2})]^{-1}\{f(-\tfrac{3}{2}, -\tfrac{3}{2})(\Phi_{-3/2,-3/2})_{00} + f(-\tfrac{3}{2}, \tfrac{1}{2})(\Phi_{-3/2,1/2})_{00} + f(\tfrac{1}{2}, -\tfrac{3}{2})(\Phi_{1/2,-3/2})_{00} + f(\tfrac{1}{2}, \tfrac{1}{2})(\Phi_{1/2,1/2})_{00} - 1\}, \tag{27d}$$

and for $(p, q) \neq (\pm\frac{1}{2}, \pm\frac{1}{2})$,

$$(\Phi_{p,q})_{00} = [g(p, q)]^{-1}\{f(p-1, q-1)(\Phi_{p-1,q-1})_{00} + f(p-1, q+1)(\Phi_{p-1,q+1})_{00} + f(p+1, q-1)(\Phi_{p+1,q-1})_{00} + f(p+1, q+1)(\Phi_{p+1,q+1})_{00}\}. \tag{27e}$$

Equations (27a)–(27e) are in a form that appears reasonably well suited to solution by relaxation techniques. We have found, however, that application of straightforward relaxation methods does not seem to yield a satisfactory result. We suspect that this is related to one or both of the following facts about this set of equations. (1) The coefficients of these simultaneous equations are such that if $\{(\Phi_{p,q})_{00}\}$ is a solution set, then so is $\{(\Phi_{p,q})_{00} + C\}$ where $C$ is any constant. (2) There are actually two sets of equations that can be solved separately. The two sets consist of those $\Phi_{p,q}$ for which $p+q$ is odd and those for which it is even. Each set can be solved independently, and then different constants, $C_{odd}$ and $C_{even}$, can be added to the sets of solution values. What this means is that the equations as written do not have any continuity between horizontally or vertically displaced adjacent elements, only between diagonally displaced adjacent elements.

To avoid these problems, to get the equations in a form where relaxation methods could be successfully applied, and incidentally to significantly reduce the size of the computational task, we have taken note of the expected symmetry of the solution to set up a reduced-size problem.

Our approach is to argue that by virtue of the symmetry of the problem,

$$\Phi_{p,q} = \Phi_{p,-q}\,. \tag{28}$$

Also, if we restrict attention to a solution in which the average value of $\Phi_{p,q}$ is zero (this restricts our choice of $C$, the additive constant), then

$$\Phi_{p,q} = -\Phi_{-p,q}\,. \tag{29}$$

TABLE I. Mean-square noise coefficient for fitting a wave-front distortion estimate to a set of noisy phase-difference measurements. The significance of the noise coefficient, $\mathfrak{N}$, is expressed by Eq. (25).

| $K$ | $\mathfrak{N}$ | $N^2$ [a] | $(2K)^2$ [b] |
|---|---|---|---|
| 2 | 0.4900 | 9 | 16 |
| 3 | 0.5832 | 25 | 36 |
| 4 | 0.6398 | 49 | 64 |
| 5 | 0.6810 | 81 | 100 |
| 6 | 0.7136 | 121 | 144 |
| 7 | 0.7405 | 169 | 196 |
| 8 | 0.7634 | 225 | 256 |
| 9 | 0.7834 | 289 | 324 |
| 10 | 0.8012 | 361 | 400 |
| 15 | 0.8684 | 841 | 900 |
| 20 | 0.9151 | 1521 | 1600 |

[a] $N^2 = (2K-1)^2$ is the number of measurement elements in the array. There are $2N^2$ phase-difference measurements made.
[b] $(2K)^2$ is the number of points at which the wave-front distortion is calculated.

This means that where previously we had to consider $(2n+2)^2$ unknowns for a $(2n+1)^2$ set of phase-difference measurement elements, we now only have to consider $(n+1)^2$ unknowns. This greatly reduces the size of our computational load. We see now that we will have basically different equations for the case of $p=\frac{1}{2}$ and/or $q=\frac{1}{2}$ than for other cases, since a quantity such as $\Phi_{p-1,q-1}$ on the right-hand side of Eq. (27e) would be replaced by $-\Phi_{p,q-1}$ for $p=\frac{1}{2}$, or $\Phi_{p-1,q}$ for $q=\frac{1}{2}$, with corresponding corrections for Eqs. (27a)–(27d). This, incidentally, would eliminate the possibility of separating the problem into two independent sets of simultaneous equations, thus insuring continuity in our solution set.

Making use of Eqs. (28) and (29) to modify Eq. (27a), we get

$$(\Phi_{1/2,1/2})_{00} = [g(\tfrac{1}{2},\tfrac{1}{2})]^{-1}\{-f(\tfrac{1}{2},\tfrac{1}{2})(\Phi_{1/2,1/2})_{00} - f(\tfrac{1}{2},\tfrac{3}{2})(\Phi_{1/2,3/2})_{00} + f(\tfrac{3}{2},\tfrac{1}{2})(\Phi_{3/2,1/2})_{00} + f(\tfrac{3}{2},\tfrac{3}{2})(\Phi_{3/2,3/2})_{00} + 1\}. \quad (30)$$

Equation (30) can be further simplified to take the form

$$(\Phi_{1/2,1/2})_{00} = [g(\tfrac{1}{2},\tfrac{1}{2}) + f(\tfrac{1}{2},\tfrac{1}{2})]^{-1}\{-f(\tfrac{1}{2},\tfrac{3}{2})(\Phi_{1/2,3/2})_{00} + f(\tfrac{3}{2},\tfrac{1}{2})(\Phi_{3/2,1/2})_{00} + f(\tfrac{3}{2},\tfrac{3}{2})(\Phi_{3/2,3/2})_{00} + 1\}. \quad (31a)$$

We note that there is no need to seek modified versions of Eqs. (27b), (27c), or (27d) since the same information can be obtained from the solution of Eq. (31a), by use of Eqs. (28) and (29). Equation (27e) is subdivided into the following three forms accordingly as $p=\frac{1}{2}$ or $q=\frac{1}{2}$, or neither equals one-half. We get

$$(\Phi_{1/2,q})_{00} = [g(\tfrac{1}{2},q)]^{-1}\{-f(\tfrac{1}{2},q-1)(\Phi_{1/2,q-1})_{00} - f(\tfrac{1}{2},q+1)(\Phi_{1/2,q+1})_{00} + f(\tfrac{3}{2},q-1)(\Phi_{3/2,q-1})_{00} + f(\tfrac{3}{2},q+1)(\Phi_{3/2,q+1})_{00}\} \quad (q>\tfrac{1}{2}), \quad (31b)$$

$$(\Phi_{p,1/2})_{00} = [g(p,\tfrac{1}{2})]^{-1}\{f(p-1,\tfrac{1}{2})(\Phi_{p-1,1/2})_{00} + f(p-1,\tfrac{3}{2})(\Phi_{p-1,3/2})_{00} + f(p+1,\tfrac{1}{2})(\Phi_{p+1,1/2})_{00} + f(p+1,\tfrac{3}{2})(\Phi_{p+1,3/2})_{00}\} \quad (p>\tfrac{1}{2}), \quad (31c)$$

and, of course,

$$(\Phi_{p,q})_{00} = [g(p,q)]^{-1}\{f(p-1,q-1)(\Phi_{p-1,q-1})_{00} + f(p-1,q+1)(\Phi_{p-1,q+1})_{00} + f(p+1,q-1)(\Phi_{p+1,q-1})_{00} + f(p+1,q+1)(\Phi_{p+1,q+1})_{00}\} \quad (p>\tfrac{1}{2} \text{ and } q>\tfrac{1}{2}). \quad (31d)$$

In preparing a computer program to solve for $(\Phi_{p,q})_{00}$ and evaluate $\mathfrak{N}$, we have considered a system with $(2K-1)^2$ measurement elements and $(2K)^2$ positions at which the phase estimates had to be formed. We therefore had to solve for $K^2$ unknowns. [Note: Hereafter we use $K$ to denote $(n+1)$, i.e., $K=n+1$.] The method of solution is a straightforward relaxation procedure with a full new solution set obtained in each iteration from the previous solution set before any of the previous solution set elements are replaced. We have carried out the solution process for $K=2, 3, 4, \ldots, 10$ and for $K=15$ and 20. To be sure that our solution had converged, the program went through $\alpha K$ iterations where $\alpha=10$ for $K=2, 3, 4, 5$, and 6, $\alpha=20$ for $K=7$, 8, 9, and 10, and $\alpha=30$ for $K=15$ and 20. The program calculated and printed out the current estimate of $\mathfrak{N}$ every tenth iteration. We were able to see that the value of $\mathfrak{N}$ had stabilized to better than four significant figures in all cases well before we reached the $\alpha K$th iteration. After the $\alpha K$th iteration, the program would print out the solution set, or rather the positive $p$, positive $q$ quadrant of the set. Then it would go through one more iteration, and print out the next value of $\mathfrak{N}$ and the updated solution set. In no case did we find any suggestion in examining any of the printouts that the iteration procedure had not fully converged, at least to within the limits of accuracy imposed by the six-significant-figure nature of the computations. The results obtained appear to be quite reliable for our purposes.

## DISCUSSION OF RESULTS

The values of $\mathfrak{N}$ calculated by the program are listed in Table I. If we examine the data in Table I, we find that it can be quite accurately estimated by the function

$$\mathfrak{N} = 0.1603\ln(K-\tfrac{1}{2}) + 0.4390 \quad (K \geq 2). \quad (32)$$

We recall that we are dealing with a system in which there are $N^2$ measurement elements (and $2N^2$ phase-difference measurements possible) and $(N+1)^2$ points at which the phase is to be estimated. Obviously, then, we can equate $K$ with the measurement array size by the formula

$$N^2 = (2K-1)^2. \quad (33)$$

If we make use of Eq. (33) to eliminate $K$ in Eq. (32), we get

$$\mathfrak{N} \approx 0.08015\ln(N^2) + 0.3279. \quad (34)$$

If we substitute this result into Eq. (25), we see that the mean-square phase error in estimating the wave-front distortion over a square aperture of $N^2$ measurement cells will be

$$\langle(\delta\Phi)^2\rangle = 0.6558[1 + 0.2444\ln(N^2)]\sigma_{pd}^2, \quad (35)$$

where $\sigma_{pd}^2$ is the mean-square error expected in measuring the phase difference in one direction across a

single measurement cell. (It is interesting to note that the very simple equation $\langle(\delta\Phi)^2\rangle \approx \frac{2}{3}[1 + \frac{1}{4}\ln(N^2)]\sigma_{pd}^2$ yields results which appear to be in error by no more than about 3%.) As can be seen from Eq. (35), the mean-square error in the phase estimate is virtually independent of the size of the array. Apparently the noise effects are very much a local matter and the presence or absence of boundaries to the array in the vicinity of each noisy measurement has very little effect on the way it introduces error into the wave-front estimate.

[1]It should be noted that while there has been an extensive body of work, and some previously published results pertaining to adaptive optics design and wave-front distortion sensing, a few examples of which we list in the references below, almost none of this work considered the approach of measuring an array of phase differences and thus had to consider the data fitting problem that we consider here.

T. R. O'Meara, U. S. Patent 3,764,213, "Return Wave Phase Control Adaptive Array," (9 October 1973).

[3]J. W. Hardy, J. Feinleib, and J. Wyant, "Real Time Phase Correction of Optical Imaging Systems," in the *Digest of Technical Papers of the OSA Topical Meeting on Optical Propagation Through Turbulence*, 9–11 July 1974, Univ. of Colorado, Boulder, Colo. (This paper considers use of a shearing interferometer to measure phase differences, but there is no discussion of the wave-front fitting problem.)

[4]W. T. Cathey, C. L. Hayes, W. C. Davis, and V. F. Pizzurro, "Compensation for Atmospheric Phase Effects At 10.6 $\mu$," Appl. Opt. 9, 701 (1970).

[5]R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," J. Opt. Soc. Am. 64, 1200–1210 (1974).

[6]C. A. Primmerman and D. G. Fouche, "Thermal-Blooming Compensation: Experimental Observation Using a Deformable Mirror System," Appl. Opt. 15, 900 (1976).

[7]W. B. Bridges, P. T. Brunner, S. P. Lazzara, T. A. Nussmeier, T. R. O'Meara, J. A. Sanguinet, and W. P. Brown, Jr., "Coherent Optical Adaptive Techniques," Appl. Opt. 13, 291 (1974).

[3]J. E. Pearson, "Atmospheric Turbulence Compensation Using Coherent Optical Adaptive Techniques," Appl. Opt. 15, 662 (1976).

# APPENDIX G

Reprinted from *Journal of the Optical Society of America,* Vol. 67(3), pp. 375-378
(March 1977).

# Wave-front reconstruction for compensated imaging

Richard H. Hudgin
*Itek Corporation, Lexington, Massachusetts 02137*
(Received 13 August 1976)

A critical component in a compensated imaging (CI) system is the wave-front sensor which measures the residual distortion of the wave front after reflecting off the active mirror. The sensor produces estimates of wave-front slopes or phase difference across the aperture. For many applications, the phase differences or slopes are not the most convenient form of data for processing or control, and they must be converted to absolute wave-front phases. This paper analyzes the conversion from phase differences to phases and derives the optimal linear estimator in terms of least noise propagation. Some remarks concerning hardware implementation are also made.

## I. INTRODUCTION

A critical component in a compensated imaging (CI) system is the wave-front sensor which measures the residual distortion of the wave front after reflecting off the active mirror. The sensor produces estimates of wave-front slopes or phase difference across the aperture. For many applications, the phase differences or slopes are not the most convenient form of data for processing or control, and they must be converted to absolute wave-front phases. This paper analyzes the conversion from phase differences to phases and derives the optimal linear estimator in terms of least noise propagation. Some remarks concerning hardware implementation are also made.[1-10]

## II. SENSOR MODEL

For a wave-front sensor divided into many subapertures, a set of noisy phase-difference estimates are produced which measure the phase differences between adjacent subapertures. For a two-dimensional array of subapertures, there are four phase differences connecting one subaperture to its neighbors. Let $\phi_{jk}$ be the two-dimensional array of subaperture phases where $j$, $k = 1, \ldots, N$, and let the connecting phase difference be defined by[11]

$$S^1_{jk} = \phi_{jk} - \phi_{j+1,k}\,, \tag{1}$$

$$S^2_{jk} = \phi_{jk} - \phi_{j,k+1}\,. \tag{2}$$

Then an $N \times N$ square array of phase points has $2N(N-1)$ connecting phase differences.

For discrete subapertures, the errors on the phase estimates due to sensor noise or photon noise is well described as uncorrelated between subapertures. Thus the noise on the phase difference measurement $S^l_{jk}$ will be denoted $n^l_{jk}$ and will be assumed to have statistics[12]

$$\langle n^l_{jk} n^{l'}_{j'k'} \rangle = n_0 \delta_{jj'} \delta_{kk'} \delta_{ll'}\,. \tag{3}$$

The actual measurements emerging from the wave-front sensor will be denoted by $\tilde{S}^l_{jk}$ and are the phase differences plus the noise error:

$$\tilde{S}^l_{jk} = S^l_{jk} + n^l_{jk}\,. \tag{4}$$

## III. RECONSTRUCTOR MODEL

The reconstructor is assumed to take the phase-difference estimates $\tilde{S}^l_{jk}$ and combine them linearly to es-

timate the wave-front phases $\phi_{jk}$. Thus the most general form for the reconstructor algorithm is

$$\tilde{\phi}_{jk} = \sum_{lmn} b^{*l}_{jkmn} \tilde{S}^l_{mn} , \tag{5}$$

where the * differentiates this $b$ from a later symbol.

The error in this estimator is

$$\epsilon_{jk} = \tilde{\phi}_{jk} - \phi_{jk} = \left( \sum_{lmn} b^{*l}_{jkmn} S^l_{mn} - \phi_{jk} \right) + \sum_{lmn} b^{*l}_{jkmn} n^l_{mn} . \tag{6}$$

We want to consider only estimators which give a perfect reconstruction when there is no noise (instantaneously unbiased estimators). Thus one condition on the $b^{*l}_{jkmn}$ is

$$\phi_{jk} = \sum_{lmn} b^{*l}_{jkmn} S^l_{mn} . \tag{7}$$

When there is measurement error, then the $b^{*l}_{jkmn}$ should be chosen to minimize the error effects in the $\tilde{\phi}_{jk}$. Here the choice is to minimize the mean-square error in $\tilde{\phi}_{jk}$, i.e., to minimize

$$\left\langle \sum_{jk} \epsilon^2_{jk} \right\rangle = n_0 \sum_{jklmn} (b^{*l}_{jkmn})^2 , \tag{8}$$

where $\langle\ \rangle$ indicates the statistical expectation value.

## IV. MINIMIZATION

The general task is now to minimize $\epsilon^2$ where

$$\epsilon^2 = \left\langle \sum_{jk} \epsilon^2_{jk} \right\rangle = n_0 \sum_{\substack{lmn \\ jk}} (b^{*l}_{jkmn})^2 \tag{9}$$

subject to the constraint that

$$\phi_{jk} = \sum_{lmn} b^{*l}_{jkmn} S^l_{mn} \tag{10}$$

for any set of $\phi_{jk}$.

For large arrays we assume that edge effects are small and that $b^{*l}_{jkmn}$ can be written as a function of $j-m$ and $k-n$ (translation invariance of $b^{*l}_{jkmn}$). Thus the solution to be derived applies to large arrays and will be perhaps suboptimal for finite arrays. It has also been demonstrated that the solution to be derived in the large array limit applies to a 2×2 array (small array limit). Thus both limits are correct and the deviation for intermediate arrays is likely small if nonzero. However, the question of optimality for a finite array remains open.

Under this large array assumption, Eq. (10) for the unbiased estimator becomes

$$\phi_{jk} = \sum_{lmn} b^l_{j-m,k-n} S^l_{mn} , \tag{11}$$

which is a convolution.

Now we go to Fourier transform space since there convolution becomes simple multiplication. A $\hat{}$ will denote a Fourier transform:

$$\hat{\phi}_{jk} = \sum_l \hat{b}^l_{jk} \hat{S}^l_{jk} , \tag{12}$$

where the different Fourier functions are defined by

$$\hat{\phi}_{jk} = \sum_{lm} \phi_{lm} e^{(2\pi i/N)(lj+mk)} , \tag{13}$$

$$\hat{b}^n_{jk} = \sum_{lm} b^n_{lm} e^{(2\pi i/N)(lj+mk)} , \tag{14}$$

$$\hat{S}^n_{jk} = \sum_{lm} S^n_{lm} e^{(2\pi i/N)(lj+mk)} . \tag{15}$$

The minimization condition in Eq. (9) becomes (in Fourier space) minimizing in the following quantity:

$$\delta^2 = \sum_{jkn} (\hat{b}^n_{jk})^2 . \tag{16}$$

Now $S^l_{jk}$ is the difference between two $\phi_{mn}$ values defined by

$$S^1_{jk} = \phi_{jk} - \phi_{j+1,k} , \tag{17}$$

$$S^2_{jk} = \phi_{jk} - \phi_{j,k+1} . \tag{18}$$

In terms of Fourier components this is equivalent to

$$\hat{S}^1_{jk} = (1 - e^{2\pi ij/N}) \hat{\phi}_{jk} , \tag{19}$$

$$\hat{S}^2_{jk} = (1 - e^{2\pi ik/N}) \hat{\phi}_{jk} . \tag{20}$$

Substituting these expressions for $\hat{S}^l_{jk}$ in Eq. (12) gives a simple form of the unbiased condition in Fourier space:

$$\hat{\phi}_{jk} = [\hat{b}^1_{jk}(1 - e^{2\pi ij/N}) + \hat{b}^2_{jk}(1 - e^{2\pi ik/N})]\, \hat{\phi}_{jk} . \tag{21}$$

Since this equation is to be satisfied for any $\phi_{jk}$ distribution, the condition on $\hat{b}^l_{jk}$ is

$$\hat{b}^1_{jk}(1 - e^{2\pi ij/N}) + \hat{b}^2_{jk}(1 - e^{2\pi ik/N}) = 1 . \tag{22}$$

The minimization condition in Eq. (16) does not couple different Fourier components, so it implies for $\hat{b}^l_{jk}$ that the following expression is minimized for each $(jk)$ value subject to the constraint of Eq. (22):

$$\delta^2_{jk} = (\hat{b}^1_{jk})^2 + (\hat{b}^2_{jk})^2 . \tag{23}$$

The problem is now a simple application of variational calculus with the result that the optimum choice for $\hat{b}^l_{jk}$ is

$$\hat{b}^1_{jk} = \frac{(1 - e^{-2\pi ij/N})}{4 - e^{2\pi ij/N} - e^{-2\pi ij/N} - e^{2\pi ik/N} - e^{-2\pi ik/N}} , \tag{24}$$

$$\hat{b}^2_{jk} = \frac{(1 - e^{-2\pi ik/N})}{4 - e^{2\pi ij/N} - e^{-2\pi ij/N} - e^{2\pi ik/N} - e^{-2\pi ik/N}} . \tag{25}$$

## V. PHYSICAL INTERPRETATION

This result for the optimum $\hat{b}^l_{jk}$ has a very simple interpretation when transformed back from Fourier space. In Fourier space the estimated value of $\hat{\phi}_{jk}$ is found by substituting Eqs. (24) and (25) into Eq. (12) to get

$$\hat{\phi}_{jk} = \frac{\hat{S}^1_{jk}(1 - e^{-2\pi ij/N}) + \hat{S}^2_{jk}(1 - e^{-2\pi ik/N})}{4 - e^{2\pi ij/N} - e^{-2\pi ij/N} - e^{2\pi ik/N} - e^{-2\pi ik/N}} . \tag{26}$$

Now multiply this equation by the denominator on the right side and realize that multiplying by an exponential in Fourier space simply translates the function in non-Fourier space. The result is

$$\hat{\phi}_{jk}[4 - e^{2\pi ij/N} - e^{-2\pi ij/N} - e^{2\pi ik/N} - e^{-2\pi ik/N}] = \hat{S}^1_{jk}(1 - e^{-2\pi ij/N}) + \hat{S}^2_{jk}(1 - e^{-2\pi ik/N}) , \tag{27}$$

| N | $\Sigma b^{2\ell}_{jkmn}$ |
|---|---|
| 3 | 0.683 |
| 5 | 0.726 |
| 7 | 0.754 |
| 11 | 0.807 |
| 13 | 0.825 |
| 20 | 0.870 |

FIG. 1. Noise coefficient for various square arrays of size $N \times N$.

or in non-Fourier space this becomes

$$4\phi_{jk} - (\phi_{j+1,k} + \phi_{j-1,k} + \phi_{j,k+1} + \phi_{j,k-1}) = S^1_{jk} - S^1_{j-1,k} + S^2_{jk} - S^2_{j,k-1} \, . \tag{28}$$

Solving for $\phi_{jk}$ gives

$$\phi_{jk} = \tfrac{1}{4}(\phi_{j+1,k} + \phi_{j-1,k} + \phi_{j,k+1} + \phi_{j,k-1}) + \tfrac{1}{4}(S^1_{jk} - S^1_{j-1,k} + S^2_{jk} - S^2_{j,k-1}) \, . \tag{29}$$

Now $S^1_{jk}$ is the measured phase difference connecting $\phi_{jk}$ and $\phi_{j+1,k}$, and $S^2_{jk}$ connects $\phi_{jk}$ and $\phi_{j,k+1}$ [see Eqs. (1)–(2)]. Thus Eq. (29) has the interpretation that $\phi_{jk}$ is the average of its four nearest neighbors plus the average of the four connecting phase differences (with appropriate signs in the phase differences).

## VI. RECURSIVE IMPLEMENTATION

The simple interpretation of Eq. (29) allows for a recursive implementation of Eq. (29) where the new value $\phi'_{jk}$ of $\phi_{jk}$ is defined

$$\phi'_{jk} = \tfrac{1}{4}(\phi_{j+1,k} + \phi_{j-1,k} + \phi_{j,k+1} + \phi_{j,k-1}) + \tfrac{1}{4}(S^1_{jk} - S^1_{j-1,k} + S^2_{jk} - S^2_{j,k-1}) \, . \tag{30}$$

The immediate questions are uniqueness and convergence of this recursive algorithm, and they can be demonstrated using standard linear algorithm techniques.

## VII. NOISE PROPAGATION

Now that the convergence, existence, and uniqueness of a solution have been proven for the recursive algorithm, we return to the noise analysis. The mean-square error $\epsilon^2$ on the reconstruction phases was [from Eq. (9)]

$$\epsilon^2 = n_0 \sum_{jkl\,mn} b^{2\,l}_{jkmn} \, , \tag{31}$$

where $n_0$ was the mean-square error on a single phase difference measurement. The $b^l_{jkmn}$ have been calculated using the recursive algorithm and the $\sum b^{2\,l}_{jkmn}$ is tabulated for various square arrays in Fig. 1.

Note that for the arrays considered, the error propagation is less than unity, i.e., *less* mean-square error appears on the reconstructed phases than goes in on the phase-difference measurements. A good fit to the data on the noise coefficient is

$$\sum_{jkl\,mn} b^{2\,l}_{jkmn} \simeq 0.561 + 0.103 \ln N \, , \tag{32}$$

where $N^2$ = number of phase points in the square array (see Fig. 1).

In addition, a simultation was constructed where random phase differences were generated and then used with the least-squares algorithm. The agreement on error propagation was within 5% of the analytical results in the two cases tested of $N = 13$ and $N = 20$.

## VIII. NOISE CORRELATIONS

Another aspect of noise error is its correlation across the aperture. This is significant for calculating optical transfer functions and optimal control algorithms for CI systems. Here we calculate that correlation function.

The error on the phase estimate is

$$\epsilon_{jk} = \sum_{l\,mn} b^{*\,l}_{jkmn} n^l_{mn} \simeq \sum_{l\,mn} b^l_{j-m,k-n} n^l_{mn} \, , \tag{33}$$

which gives a correlation function of

$$C_{rs} = \langle \epsilon_{j+r,k+s} \epsilon_{jk} \rangle \tag{34}$$

$$= n_0 \sum b^l_{r+m,s+n} b^l_{mn} \, . \tag{35}$$

Note that for $r = s = 0$ this becomes the noise error variance calculated in Sec. III.

The $b^l_{jk}$ have been calculated for a $14 \times 14$ array of phase points using the recursive algorithm and the noise correlation function $C_{rs}$ plotted. It is very nearly rotationally symmetric so it is plotted in Fig. 2 as a function of radial spacing $(r^2 + s^2)^{1/2}$. The correlation is seen to be quite long range, extending over roughly one-half the aperture. The correlation function is approximately expressed

$$C_{rs} = C_{00}\, e^{-5(r^2+s^2)^{1/2}/14} \, . \tag{36}$$

## IX. HARDWARE IMPLEMENTATION

The least-squares recursive algorithm has been implemented in digital and analog forms for compensated imaging systems. The algorithm requires a resistor net and $N^2$ op amps for analogue implementation and $N^2$ 16-bit adders for digital operation. Note that dividing by 4 is a bit shift not a full division. The errors

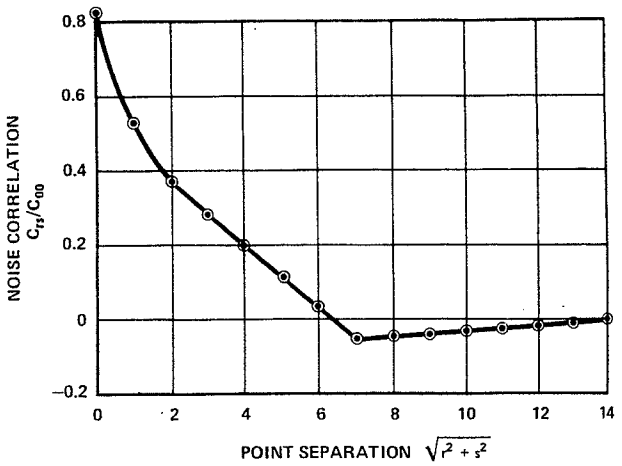



FIG. 2. Noise correlation function on the reconstructed phases.

induced by component tolerance and bit limitations have been seen to be negligible, and convergences times of around 10 $\mu$s are reasonable for special-purpose, low-power hardware.

## X. CONCLUSION

The optimum linear algorithm has been derived which converts noisy phase differences to absolute phase estimates for a wave-front sensor. The algorithm has been given a simple, recursive form, and convergence and uniqueness have been shown. The noise propagation for this algorithm depends weakly on the size of the phase array (logarithmically) and is less than unity for the cases studied. This means that there is less rms error on the reconstructed phases than on the phase-difference measurements from the sensor. This error propagation is in marked contrast to a simple summing of phase differences which can produce many times more error.

Hardware implementation, either in analog or digital form, is easily achieved and reconstruction times of 10 $\mu$s are reasonable for low-power components.


[1]J. Hardy, J. Lefebvre, and C. Koliopoulos, "Real-time atmospheric compensation," J. Opt. Soc. Am. 67, 360–369 (1977) (this issue).

[2]J. Hardy, C. Koliopoulos, and K. Bowker, "Real-time AC grating interferometer wave-front sensor," (unpublished).

[3]H. W. Babcock, Pub Astr. Soc. Pac. 65, 229 (1953).

[4]"Restoration of Atmospherically Degraded Images," Woods Hole Summer Study, July 1966. DDC Reports AD806878, AD806879, AD806880, AD680807.

[5]J. W. Hardy, J. Feinleib, and J. C. Wyant, "Real-time correction of optical imaging systems," presented at the OSA Meeting on Optical Propagation through Turbulence, Boulder, Colo. July 1974.

[6]J. C. Wyant, "White Light Extended Source Shearing Interferometer," Appl. Opt. 13, 200 (1974).

[7]Analog Data Processor, US Patent, 3,921,080 November 18, 1975, assigned to Itek Corporation, Lexington, Mass.

[8]J. Feinleib, S. G. Lipson, and P. F. Cone, "Monolithic piezoelectric mirror for wavefront correction," Appl. Phys. Lett. 25, 311 (1974).

[9]J. W. Hardy, C. L. Koliopoulous, and J. K. Bowker, "Radial Grating A. C. Interferometer" (unpublished).

[10]Freeman J. Dyson, "Photon noise and atmospheric noise in active optical systems," J. Opt. Soc. Am. 65, 551–558 (1975).


[11]When the wave front must be regarded as changing over the time interval of the sensor measurement, then $\phi_{jk}$ will be the average of the mean phase over the interval and $S^t_{jk}$ the average of the phase difference.

[12]We note that even for a Hartmann sensor (where the light from each subaperture is imaged onto a quad cell, and $x$-$y$ tilts are measured using same photons), the probability density for photon arrival for a square subaperture is of the form $P(x)$-$f(x)f(y)$ for an unresolved target. This means that for a single photon, the $x$ coordinate of arrival is uncorrelated with the $y$ coordinate of arrival, so the photon noises on the two components of tilt are uncorrelated even though using the same photons. Thus the analysis applies to this case. For a round subaperture the independence is based on a $\pm$ symmetric error distribution, but again the conclusion of uncorrelated tilt errors holds.

# APPENDIX H

Reprinted from *Journal of the Optical Society of America,* Vol. 67(3), pp. 378-382
(March 1977).

# Optimal wave-front estimation

Richard H. Hudgin
*Itek Corporation, Lexington, Massachusetts 02173*

In a compensated imaging (CI) system the wave-front distortion is changing randomly in time and the data processing must follow the changes and even, to some extent, predict them. Furthermore, the system is not infinitely fast; it has time delays and bandwidth limitations which can be quite significant. Couple to this the substantial photon noise in the wave-front measurements, and it becomes desirable to find the optimal data processing to use for system control. This paper derives the optimal linear control algorithm assuming the aberration and noise statistics are known for a compensated imaging system, and discusses the properties of this algorithm.

## I. INTRODUCTION

In a compensated imaging (CI) system the wave-front distortion is changing randomly in time and the data processing must follow the changes and even, to some extent, predict them. Furthermore, the system is not infinitely fast; it has time delays and bandwidth limitations which can be quite significant. Couple to this the substantial photon noise in the wave-front measurements, and it becomes desirable to find the optimal data processing to use for system control. This paper derives the optimal linear control algorithm, assuming the aberration and noise statistics are known for a compensated imagining system, and discusses the properties of this algorithm.

## II. SYSTEM MODEL

A CI system consists of three main components: a wave-front sensor, a data processor, and an active mirror. The wave-front sensor produces photon noisy wave-front estimates at every time interval $\delta$ and inputs them into the data processor. The processor, in turn, combines the measurements linearly and generates a predicted wave front which is sent as commands to the active mirror.

Thus we will assume a set of $N$ phase estimation points at positions $X_j : j = 1 - N$, and the wave-front sensor will generate noisy phase measurements $\bar{\phi}_0(\mathbf{x}_j, t - K\delta)$ for $K = K_0, -\infty$.

The value of $K_0$ represents the time delay of the system, measured in units of the sensor interval $\delta$. An ultimate system would have $K_0 = 1$, (i.e., the data is used right after it is measured with a mean delay of one cycle), but the data processing requirements may be too heavy for that limit to apply. Thus $K_0 = 2$ or 3 may be more meaningful for an operational system, and the performance of the estimator as a function of the delay $K_0\delta$ is studied below. For a postprocessing estimate of the wave front, all data can be used, so $K_0 = -\infty$ in that case. Actually, there will be a finite cutoff in the number of wave fronts processed due to CPU limitations even for postprocessing estimates of wavefront error. The effects of restricting the number of terms are also studied.

The data processor takes these noisy wave-front estimates $\tilde{\phi}_0(\mathbf{x}_j, t - K\delta)$ and combines them linearly with weights $A_{ijk}$ to get an estimate $\hat{\phi}(\mathbf{x}_i, t)$ for the wave-front at time $t$:

$$\hat{\phi}(\mathbf{x}_i, t) = \sum_{jk} A_{ijk}\tilde{\phi}_0(\mathbf{x}_j, t - k\delta) . \tag{1}$$

Note that the $A_{ijk}$ will usually be different for each $\mathbf{x}_i$.

The error of this estimate is the difference between the estimate $\hat{\phi}(\mathbf{x}_i, t)$ and the actual distortion $\phi_a(\mathbf{x}_i, t)$. The mean-square error $E$ will be calculated below and depends on the weights $A_{ijk}$, the wave-front statistics, the cycle time $\delta$, the time delay $K_0\delta$, and the sensor noise $n(\mathbf{x}_j, t - k\delta)$ in the measurements $\tilde{\phi}_0(\mathbf{x}_j, t - k\delta)$.

Perhaps we might mention here that several sources of error are not considered in this analysis. The error due to discretely sampling the wave front is not calcalculated, and is treated elsewhere.[12] Also, errors due to atmospheric dispersion,[13] miscellaneous system optical distortions, and isoplanatism are not considered. The final system error will be a sum of these different error terms, so the present paper is by no means a full system analysis.

## III. MEAN-SQUARE ERROR

The error variance $E$ of the estimate $\hat{\phi}$ can be calculated from the statistics of the wave-front error and the sensor noise. We choose to characterize the wave-front statistics by a structure function $D$ defined by

$$D(\mathbf{x}, t) = \langle [\phi_a(\mathbf{y}+\mathbf{x}, t'+t) - \phi_a(\mathbf{y}, t')]^2 \rangle , \tag{2}$$

where $\langle\ \rangle$ denotes statistical expectation value.

The sensor noise is assumed to have a known correlation function $C_{jklm}$ defined by

$$C_{jklm} = \langle n(\mathbf{x}_j, t - k\delta)\, n(\mathbf{x}_l, t - m\delta) \rangle . \tag{3}$$

In the case of a wave-front sensor which is a shearing interferometer or a Hartmann sensor, the function $C_{jklm}$ has been analyzed[4] and is approximately expressed as

$$C_{jklm} \simeq C_0\, e^{-s|\mathbf{x}_j - \mathbf{x}_l|/D_0}\, \delta_{km} , \tag{4}$$

where $D_0$ is the diameter of the telescope aperture and $C_0$ is the mean-square noise. Note that measurements at different times (i.e., $K \neq m$) have independent photons and thus independent noise errors.

Using these statistics in Eqs. (2) and (3), one can calculate the error variance $E$ as follows:

$$E_i = \langle [\phi_a(\mathbf{x}_i, t) - \hat{\phi}(\mathbf{x}_i, t)]^2 \rangle \tag{5}$$

$$= \left\langle \left[ \phi_a(\mathbf{x}_i, t) - \sum_{\substack{j=1\to N \\ k=k_0\to\infty}} A_{ijk}\tilde{\phi}_0(\mathbf{x}_j, t - k\delta) \right]^2 \right\rangle , \tag{6}$$

where we substituted Eq. (1) for $\hat{\phi}$.

Now express $\tilde{\phi}_0$ as $\phi_a$ plus the noise error $n$ to get

$$E_i = \left\langle \left[ \phi_a(\mathbf{x}_i, t) - \sum_{\substack{j=1\to N \\ k=k_0\to\infty}} A_{ijk}(\phi_a(\mathbf{x}_j, t - k\delta) + n(\mathbf{x}_j, t - k\delta) \right]^2 \right\rangle . \tag{7}$$

It is actually not the absolute phase that is measurable in the aperture, but relative phase. Adding a uniform phase to the aperture is not physically meaningful; thus the mean-square error $E_i$ should involve wave fronts with zero mean, and to get the physically meaningful error, we interpret $\phi_a$ as a wave front of zero mean. Subtracting this mean allows the use of the wave-front structure function in calculating $E_i$ and gives the following result:

$$E_i = \left\langle \left\{ \phi_a(\mathbf{x}_i, t) - \frac{1}{N}\sum_{l=1\to N} \phi_a(\mathbf{x}_l, t) - \sum_{\substack{j=1\to N \\ k=k_0\to\infty}} A_{ijk}[\phi_a(\mathbf{x}_j, t - k\delta) + n(\mathbf{x}_j, t - k\delta)] + \frac{1}{N}\sum_{\substack{j,\, l=1\to N \\ k=k_0\to\infty}} A_{ijk}[\phi_a(\mathbf{x}_l, t - k\delta) + n(\mathbf{x}_l, t - k\delta)] \right\}^2 \right\rangle \tag{8}$$

$$= E_{i0} + \sum_{\substack{j=1\to N \\ k=k_0\to\infty}} A_{ijk}\bar{B}_{jk} + \sum_{\substack{j,\, l=1\to N \\ k,\, m=k_0\to\infty}} A_{ijk}A_{ilm}\bar{C}_{jklm} , \tag{9}$$

where

$$E_{i0} = \frac{1}{N}\sum_{j=1\to N}\{D(\mathbf{x}_i - \mathbf{x}_j, 0) + C_{ij0j0}\} - \frac{1}{2N^2}\sum_{j,\, l=1\to N}\{D(\mathbf{x}_j - \mathbf{x}_l, 0) + C_{j0l0}\} , \tag{10}$$

$$\bar{B}_{jk} = D(\mathbf{x}_i - \mathbf{x}_j, k\delta) + C_{ikj0} - \frac{1}{N}\sum_{l=1\to N}\{D(\mathbf{x}_i - \mathbf{x}_l, k\delta) + D(\mathbf{x}_j - \mathbf{x}_l, -k\delta) + C_{ikl0} + C_{j0lk}\} + \frac{1}{N^2}\sum_{l,\, m=1\to N}\{D(\mathbf{x}_l - \mathbf{x}_m, k\delta) + C_{ikm0}\} , \tag{11}$$

$$\bar{C}_{jklm} = -\tfrac{1}{2}D[\mathbf{x}_j - \mathbf{x}_l, (m-k)\delta] - \tfrac{1}{2}C_{jklm} + \frac{1}{2N}\sum_{n=1\to N}\{D[\mathbf{x}_n - \mathbf{x}_l, (m-k)\delta] + D[\mathbf{x}_j - \mathbf{x}_n, (m-k)\delta] + C_{nklm} + C_{jknm}\} - \frac{1}{2N^2}\sum_{n,\, r=1\to N}\{D[\mathbf{x}_n - \mathbf{x}_r, (m-k)\delta] + C_{nkrm}\} . \tag{12}$$

Now Eq. (9) is a quadratic equation involving the $A_{ijk}$. The values of $A_{ijk}$ which minimize $E_i$ can be found by setting

the variational derivative of $E_i$ with respect to $A_{ijk}$ equal to zero. This produces the following set of linear, simultaneous equations which can be solved for the $A_{ijk}$:

$$2\sum_{lm} A_{ilm} C_{jklm} + B_{jk} = 0 . \quad (13)$$

## IV. TEMPORAL AVERAGING

The first attempt to use these equations was to take the past wave fronts at a single point and to estimate the correction from them. Thus only temporal weighting, not spatial weighting is involved, which implies

$$A_{ijk} = 0 \text{ for } j \neq i . \quad (14)$$

Here we set $K_0 = 2$ to represent a one cycle processing delay in the system and found the error as a function of sensor noise and $\delta$ for one, two, and three terms in the past (with optimum weights). Including more than three terms caused little improvement in the range studied. The results are presented in normalized form in Fig. 1.

We have assumed a Kolmogorov atmospheric structure function where

$$D(0, t) = d_0 t^{5/3} \text{ (waves}^2\text{)} , \quad (15)$$

and have used the noise correlation function

$$C_{jkjm} = C_0 \delta_{km} \text{ (waves}^2\text{)} , \quad (16)$$

where each wave-front measurement has its own independent noise error.

If there is no photon noise, then the error using a single term which is two cycles old is $d_0(2\delta)^{5/3}$. This is the $y$ intercept of the "one wave-front" curve. Using more terms is always better, but only slightly until the photon noise gets very large compared to the time delay error for one cycle, i.e.,

$$C_0 \gg d_0 \delta^{5/3} . \quad (17)$$

When the photon noise is that large, then a larger $\delta$ can be used without significant degradation. Thus a two-term predictor is nearly optimal when spatial correlations are ignored.

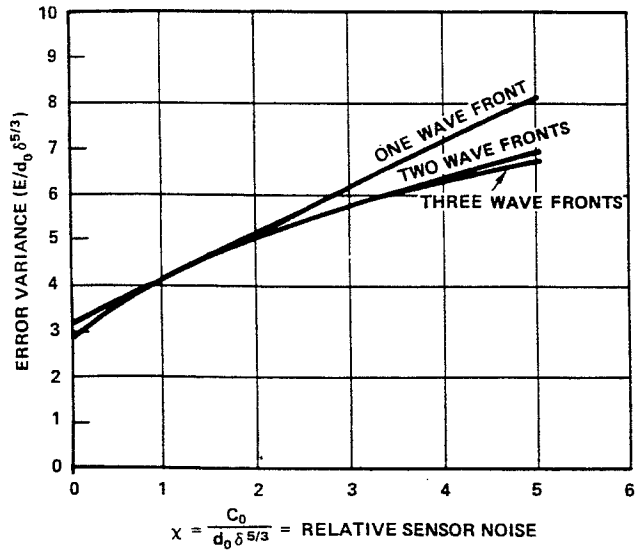



FIG. 1. Dependence of the estimation error on sensor noise and number of terms for temporal averaging.

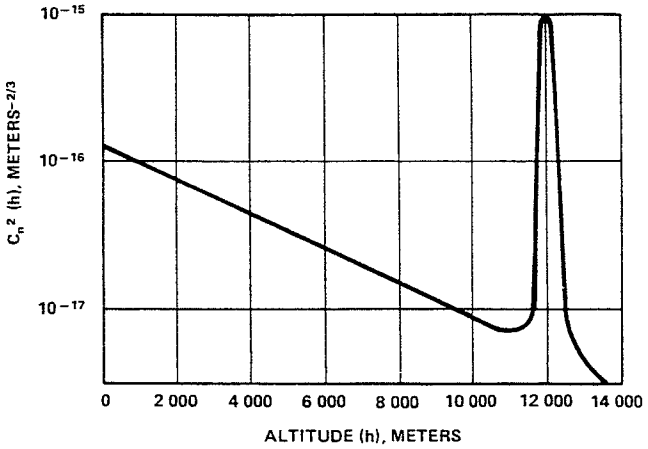



FIG. 2. Atmospheric turbulence distribution used to evaluate the estimator gives $r_0 = 10$ c (equal turbulence in tropopause peak and the exponential).

## V. SPACE-TIME WEIGHTING

The residual error of the optimum space-time estimator was studied next. We chose an array of phase points that was $21 \times 21$ and used 40 wave fronts in the past. The 200 phase points with the largest correlation to the point to be estimated are chosen out of the $21 \times 21 \times 40 = 18\,081$ candidates. Then the 200 optimal weights are calculated. The resulting mean-square error is felt to be a close approximation to the minimum possible error.

When the effects of fewer terms were explored, the $N$ terms with the largest weights among the 200 are used. The optimum weights are recalculated for these $N$ terms, and then the mean-square error is derived.

The dependence of the error on position in the $21 \times 21$ aperture was also studied as well as the selection pattern of the terms.

The wave-front statistics were derived from a 20-layer atmosphere with an exponentially decaying low-level turbulence and a strong tropopause layer of turbulence. The turbulence strength $C_n^2$ is plotted as a function of height in Fig. 2, and corresponds to a coherence length of $r_0 = 10$ cm (i.e., 1 arc sec seeing). The structure function used was

$$D(\mathbf{x}, t) = \frac{2.91}{\lambda^2} \int dh\, C_n^2(h) \left| \mathbf{x} - h\omega t \right|^{5/3} \text{ (waves}^2\text{)} , \quad (18)$$

where $\omega$ is the angular slew velocity and $\lambda$ is the mean wavelength of light. This corresponds to a Kolmogorov type of turbulence which moves due to the slewing of the line of sight through the atmosphere.

The noise correlation function used was the one in Eq. (4).

## VI. OPTIMUM SPACE-TIME ESTIMATION

The results of the 200-term estimator are shown in Fig. 3 for various levels of sensor noise and system

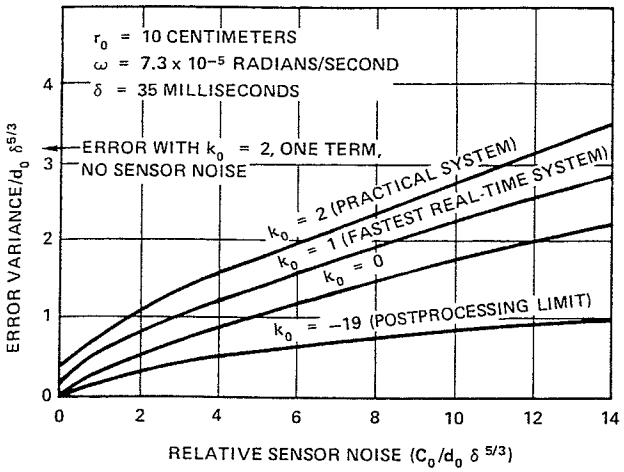



FIG. 3. Dependence of estimation error (100 terms) on system speed and sensor noise.

delay. The slew rate corresponds to the rotational velocity of the earth, and the sensor interval is $\delta = 35$ ms. The error increases as sensor noise increases and for a real-time system does not go to zero (even without sensor noise) due to a time delay of at least one-cycle $\delta$ between when the data is measured and when it is used. This delay error is gone as soon as the postprocessing region is entered ($K_0 \leq 0$). There the wave front to be estimated is observed, and thus the estimation error vanishes when sensor noise vanishes.

If there were no sensor noise, the error for a practical system with $K_0 = 2$ (i.e., one cycle interval $\delta$ for data processing) would be $d_0(2\delta)^{5/3}$ if only the last

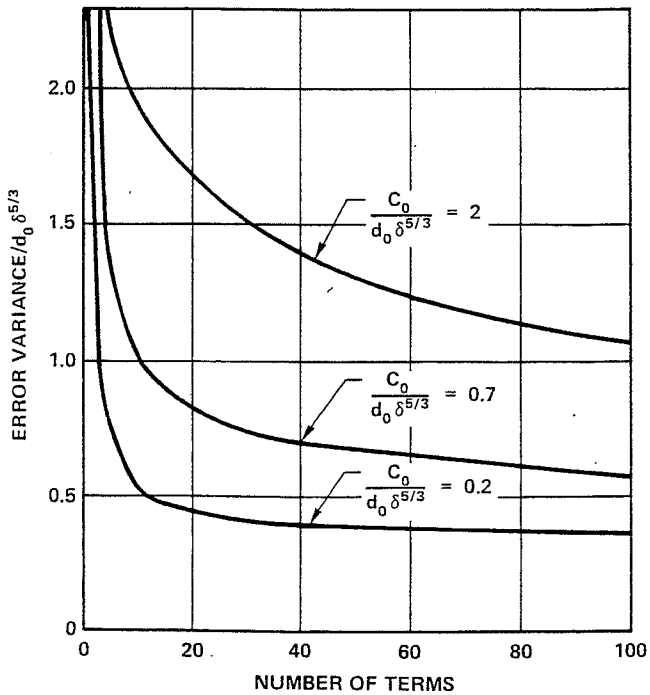



FIG. 4. Predictor error variance vs relative sensor noise and number of terms ($K_0 = 2$).

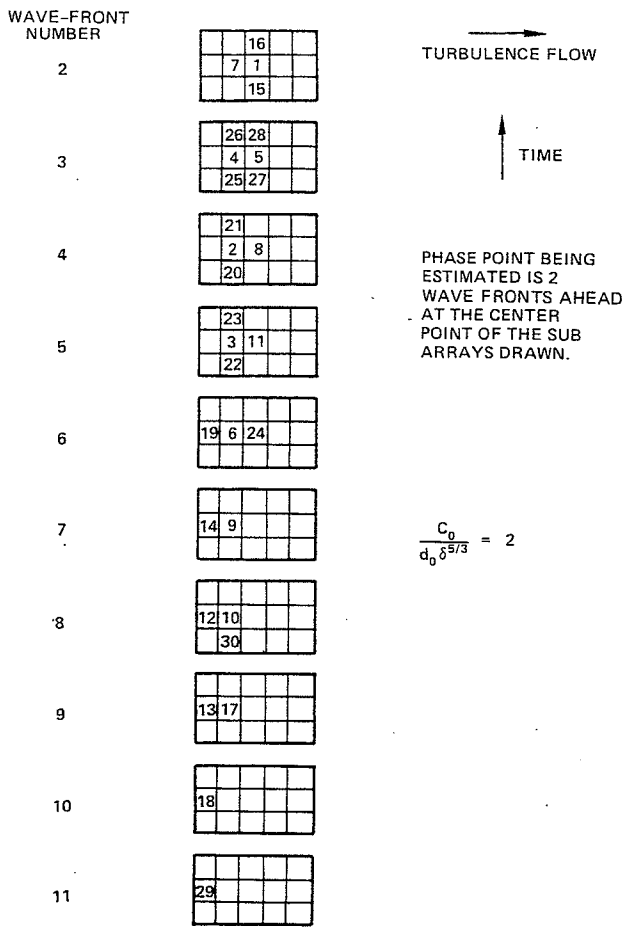



FIG. 5. Term ordering by weight for the optimal estimator.

measurement were used in the estimator. The reduction in error by a factor of 10 from the one-term estimator to the optimal estimator indicates the importance of using more than one term.

*A posteriori* estimates (here approximated by $K_0 = -19$ which used 19 past and 20 future wave fronts to estimate a current wave front) are substantially better than the real-time estimates. They can be used for possible postprocessing applications

## VII. DEPENDENCE ON THE NUMBER OF TERMS

In Fig. 3 the contrast for $K_0 = 2$ between one term and 200 terms was considerable. In Fig. 4, the dependence of the estimator error on the number of terms for a practical system ($K_0 = 2$) is shown. For sensor noise smaller than or equal to the error due to the system time delays, roughly 20 terms will give nearly optimum performance. As the sensor noise increases, more averaging is required so that more terms are useful. If sensor noise increased beyond twice the system delay error, then using a longer sensor integration time is probably preferable to using more terms.

## VIII. TERM SELECTION

In Fig. 5 the order of the first 30 terms is tabulated by the magnitudes of their weights. The selection pat-

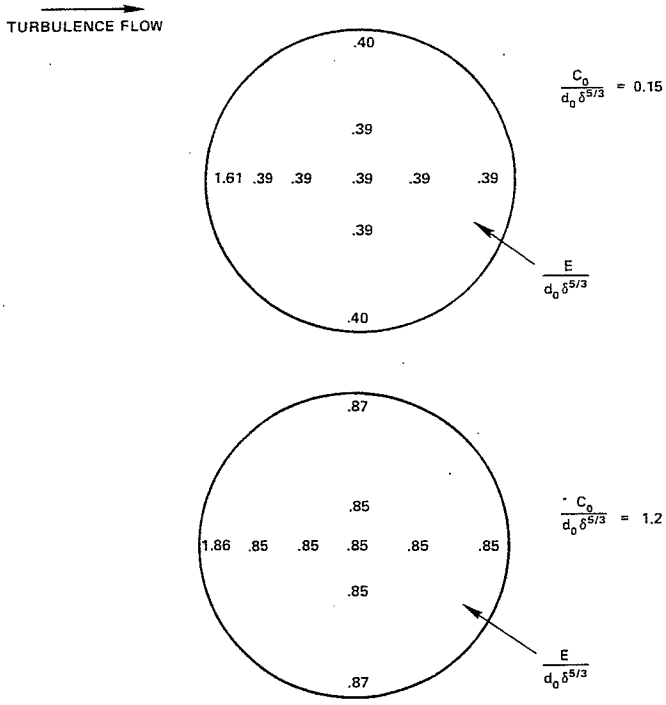



FIG. 6. Estimation error as a function of position in the aperture.

tern follows the flow the turbulence back in time and spreads out only very slightly orthogonal to the turbulence flow. The first wave front available is two behind the one being estimated for a $K_0 = 2$ system.

## IX. SPATIAL DEPENDENCE OF THE ESTIMATOR

One question to consider in designing a CI system is how the error varies across the aperture. If the edges of the wave front are significantly worse than the interior, then masking off the aperture before imagining may improve the system performance.

The numerical analysis has been performed for a variety of locations in an aperture with a $21 \times 21$ point grid with the results presented in Fig. 6. There is a slight increase in estimation error near the edges of the aperture, but it is only significant right at the leading edge when the turbulence is just entering. Consequently, only the leading edge would be a candidate for masking off. The results are presented for two levels of sensor noise with the same general result.

## X. CONCLUSION

The optimum linear estimator for a compensated imagining system with sensor noise has been developed and applied to various cases. In general, a substantial improvement results from using more than the last sensor measurement. Spatial correlations in the wave front are seen to be significant for Kolmogorov turbulence and the estimation is very poor when they are ignored.

For finite apertures, the accuracy of the estimator is nearly independent of position except at the leading edge of the aperture where the new turbulence is appearing. It may be desirable to mask off the leading edge in a practical CI system.

The number of terms required to get nearly optimal performance in the estimator increases as the sensor noise increases, but for sensor noise less than or equal to the error caused by system time delay, 20 terms is nearly optimal.

Postprocessing estimates of the wave front are substantially more accurate than real-time estimates which may be useful in image enhancement.


[1]J. Hardy, J. Lefebvre, and C. Koliopoulos, "Real-time atmospheric compensation," J. Opt. Soc. Am. 67, 360–369 (1977) (this issue).

[2]J. Hardy, C. Koliopoulos, and K. Bowker, "Real-time AC grating interferometer wavefront sensor," (unpublished).

[3]R. Hudgin, "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am. 67, 375–378 (1977) (previous paper).

[4]H. W. Babcock, Pub Astr. Soc. Pac. 65, 229 (1953).

[5]"Restoration of Atmospherically Degraded Images," Woods Hole Summer Study, July 1966. DDC Reports AD806878, AD806879, AD806880, AD680807.

[6]J. W. Hardy, J. Feinleib, and J. C. Wyant, "Real-Time Correction of Optical Imaging Systems," presented at the OSA Meeting on Optical Propagation through Turbulence, Boulder, Colo. July 1974.

[7]J. C. Wyant, "White Light Extended Source Shearing Interferometer," Appl. Opt. 13, 200 (1974).

[8]Analog Data Processor, U. S. Patent, 3, 921,080 November 18, 1975, Assigned to Itek Corporation, Lexington, Mass.

[9]J. Feinleib, S. G. Lipson, and P. F. Cone, "Monolithic Piezoelectric mirror for wavefront correction," Appl. Phys. Lett. 25, 311 (1974).

[10]J. W. Hardy, C. L. Koliopoulos, and J. K. Bowker," Radial Grating A. C. Interferometer," (unpublished).

[11]Freeman J. Dyson, "Photon noise and atmospheric noise in active optical systems," J. Opt. Soc. Am. 65, 551–558 (1975).

[12]R. Hudgin, "Wave-front compensation error due to finite corrector-element size," J. Opt. Soc. Am. 67, 393–395 (1977) (this issue).

[13]E. Wallner, "Minimizing atmospheric dispersion effects in compensated imaging," J. Opt. Soc. Am. 67, 407–409 (1977) (this issue).

# APPENDIX I

multilayer antireflection coatings," J. Opt. Soc. Am. **52,** 431–436 (1962).
[3]Alfred Thelen, "Equivalent layers in multilayer filters," J. Opt. Soc. Am. **56,** 1533–1538 (1966).
[4]Curtis Ufford and Philip Baumeister, "Graphical aids in the use of equivalent index in multilayer-filter design," J. Opt. Soc. Am. **64,** 329–334 (1974).
[5]K. C. Park, "The Extreme Values of Reflectivity and the Conditions for Zero Reflection from Thin Dielectric Films on Metal," Appl. Opt. **3,** 877–881 (1964).
[6]P. A. Young and W. G. Thege, "Two-layer laser antireflection coatings," J. Phys. D 4, 64–71 (1971).
[7]J. E. Rudisill, M. Braunstein, and A. J. Braunstein, Appl. Opt. **13,** 2075–2078 (1974).
[8]A. D. Baer, "Design of three-layer antireflection coatings," Nat. Bur. Stand. (U.S.), Spec. Publ. **462,** 221–229 (1976).

# Phase estimates from slope-type wave-front sensors

**Robert J. Noll**
*The Perkin-Elmer Corporation, Norwalk, Connecticut 06856*
(Received 17 June 1977)

In a recent issue of this journal Fried and Hudgin[1,2] show that the expected mean-square wave-front error due to measurement noise over a square array of phase-difference-type wave-front measurements depends on the total number of measurement points $N$ according to $A + B \ln N$. The methods employed involve a computer least-square solution of the equation net. In this Letter, it is shown that the $\ln N$ dependence can be derived analytically, and suggested that the least-square solutions are similar to solutions of Poisson's equation with Neumann boundary conditions.

Minimizing the mean-square wave-front error for a compensated imaging system with a phase difference wave-front sensor has been shown[3] to be consistent with a wave-front corrector that solves Poisson's equation. Since only phase slopes are measured, the boundary condition involves only the normal derivative (Neumann boundary conditions). Solving Poisson's equation is a classical problem which has a wide variety of solution techniques as discussed by Morse and Feshbach,[4] among them finite difference solution methods. With Neumann boundary conditions the phase can be obtained to within an arbitrary phase provided that the boundary slopes satisfy an integral requirement like

$$\oint \mathbf{s}(\mathbf{r}) \cdot \mathbf{n}\, dl = 0, \tag{1}$$

where $\oint$ denotes integration around the boundary, $\mathbf{n}$ is a normal to the boundary, and $\mathbf{s}$ is the slope at point $\mathbf{r}$. A knowledge of the Green function for the measurement configuration is all that is required to find a solution. On a lattice denoted by $(m,l)$ the phase $\Phi$ can be obtained by

$$\Phi(m,l) = \sum_{q,t} F(q,t) g(q,t|m,l) + \sum_{B} (\mathbf{s}_B \cdot \mathbf{n}) g(q_B,t_B|m,l), \tag{2}$$

where $\mathbf{s}_B \cdot \mathbf{n}$ is the normal component of the slope measurement at the boundary and $g$ is a Green function that has a zero normal derivative at the boundary. The quantity $F$ is simply the result of taking the divergence of the slopes,

$$\nabla \cdot \mathbf{s} = -F. \tag{3}$$

The solution to the phase-difference problem therefore centers solely on finding the appropriate Green function $g$.

A simple analytic Green function exists for the infinite boundary problem. Let us consider this Green function and neglect the boundary term. For convenience let us use a continuous solution rather than the discrete form given by Eq. (2); that is,

$$\Phi(\mathbf{r}) = -\int d\mathbf{r}' [\nabla \cdot \mathbf{s}(\mathbf{r}')] g(\mathbf{r}|\mathbf{r}'). \tag{4}$$

Taking the Fourier transform of Eq. (4) and noting that the infinite boundary Green function transform takes the form

$$G(\mathbf{k}|\mathbf{k}') = (1/4\pi^2 k^2)\delta(\mathbf{k} - \mathbf{k}') \tag{5}$$

leads to the Fourier equivalent of Eq. (4),

$$\Phi(\mathbf{k}) = -i\mathbf{k} \cdot \mathbf{S}(\mathbf{k})/2\pi k^2. \tag{6}$$

This form can be used for calculating the mean-square wave-front error due to noise. Multiplying Eq. (6) by $\Phi^*(k)$ and taking an ensemble average gives

$$\langle \Phi^*(\mathbf{k})\Phi(\mathbf{k}) \rangle = \langle [\mathbf{k} \cdot \mathbf{S}^*(\mathbf{k})][\mathbf{k} \cdot \mathbf{S}(\mathbf{k})] \rangle / (2\pi k^2)^2. \tag{7}$$

Now make the usual assumption that the slope noise due to the measurement sensor is white over the system pupil and that the $x$-$y$ components are independent so that the spatial power spectral density $N_o$ of the wave-front slope measurements is defined by

$$N_o = \tfrac{1}{2} \langle \mathbf{S}^*(\mathbf{k}) \cdot \mathbf{S}(\mathbf{k}) \rangle. \tag{8}$$

The spatial power spectral density of the wave front given by Eq. (7) is thus

 0030-3941/78/6801-0139$00.50  

$$\langle \Phi^*(\mathbf{k})\Phi(\mathbf{k})\rangle = N_o/2\pi^2 k^2. \quad (9)$$

The singularity at the origin in Eq. (9) is a result of the fact that a solution can be obtained to within an arbitrary phase. Taking the constant phase over the pupil to be zero is equivalent to multiplying Eq. (9) by $1 - \Lambda_1^2(k)$, where

$$\Lambda_1^2(k) = \mathrm{sinc}^2(\pi k_x D)\,\mathrm{sinc}^2(\pi k_y D)$$

(square pupil of size $D$), (10)

$$\Lambda_1^2(k) = 4J_1^2(2\pi kR)^2/(2\pi kR)^2$$

(circular pupil of radius $R$). (11)

Thus the mean-square wave-front error due to measurement slope noise $E_N$ can be written

$$E_N = \frac{N_o}{2\pi^2}\int_{\text{Nyquist}} d\mathbf{k}\,\frac{[1-\Lambda_1^2(k)]}{k^2}, \quad (12)$$

where the limit on the integration is the spatial Nyquist frequency $K$, which is related to the total number of sampling points, $N$, in the pupil of dimension $D$ by $K = \sqrt{N}/2D$. Numerical integration of Eq. (12) yields

Square aperture: $E_N = 2N_o[0.0536 + 0.0795\ \ln N]$, (13)

Circular aperture: $E_N = 2N_o[0.0068 + 0.0796\ \ln N]$. (14)

Equation (13) should be compared with Fried's result, which in the present notation is

$$E_N = 2N_o[0.3279 + 0.0815\ \ln N]. \quad (15)$$

The only significant discrepancy between Eqs. (13) and (15) is in the constant term which we take to be due to our neglecting the boundary term in Eq. (2). It should be noted that aberration errors like tilt and astigmatism contribute only through the boundary term. This is because $\nabla \cdot \mathbf{s} = 0$ for tilt and astigmatism inside the pupil. Thus neglecting the boundary term results in omission of significant low-order aberration errors. The origin of the $\ln N$ dependence is a consequence of the $1/k^2$ power density of the wave-front phase, i.e., of the two-dimensional integration of the wave-front slopes. Put another way, white wave-front slope noise produces $1/k^2$ wave-fronts errors, which means larger errors associated with low spatial-frequency distortions.

I believe that the relaxation method used by Fried for a square lattice to obtain the mean-square wave front is similar to a method suggested by Morse and Feshbach[4] for finding the Green function. Because Eq. (2) applies for arbitrarily shaped pupils, the method should be of general usefulness in obtaining the phase from slope measurements. All that is required is the determination of the Green function corresponding to the specific geometry of the problem.


[1] D. L. Fried, J. Opt. Soc. Am. **67**, 370 (1977).
[2] R. H. Hudgin, J. Opt. Soc. Am. **67**, 375(1977).
[3] R. J. Noll, Proc. SPIE **75**, 39 (1976).
[4] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, I (McGraw-Hill, New York, 1953), p. 698.


# Brightness exponent for the periphery in the Bloch region


**Naoyuki Osaka**

*Department of Psychology, Otemon-Gakuin University, Ibaraki, Osaka 567, Japan*



(Received 11 August 1977)



Using a method of direct magnitude estimation, perceived brightness was measured in the dark-adapted eye with brief flashes of varying duration in the Bloch region (1–100 ms) and retinal lock (0°–40°) for the photopic luminance levels covering the range between 140 and 0.14 cd/m$^2$ in steps of 1 log unit. Perceived brightness increased as a function of flash luminous energy (product of luminance and duration) up to critical duration of approximately 100 ms. The brightness power exponent for brief flash was found to be 0.48 ± 0.01 in the fovea, whereas about 0.44 ± 0.01 in the periphery.

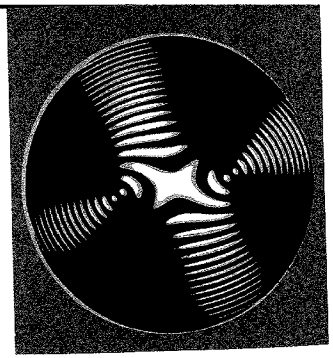

It has been established in threshold studies as Bloch's law that luminance power is integrated over time up to a critical flash duration ($t_c$) of approximately 100 ms (Bloch region), beyond which temporal summation ceases and the threshold is then defined solely in terms of flash luminance ($L$). Bloch's reciprocity law ($Lt$ = const.) and a decrease in $t_c$ with increasing $L$, in the fovea, have also been established for suprathreshold moderate luminance range in the dark-adapted eye. Perceived brightness grows as luminous-energy ($Lt$) raised to a power in the Bloch region. Brightness can be expressed as power law of the form $\psi = k\,E^{\beta \pm \text{S.E.}}$, where $\psi$, $E$, $\beta \pm$ S.E., and $k$ indicate perceived brightness, luminous energy, exponent ± standard error, and a scale factor, respectively.

In the *foveal viewing*, the brightness exponent has been found to be approximately 0.50 for brief flashes below the $t_c$, since the exponent has been estimated approximately 1½ times larger for transient-state flashes than for steady-state flashes.[1–3] If the exponent for longer flashes (out of the Bloch region) is assumed to be 0.33,[1,3] the exponent for brief flashes (within the Bloch region) may be estimated to be close to 0.50.[2,3] The study of Raab[4] gave a flash exponent of 0.5, that of Mansfield,[2] 0.50 ± 0.03, that of Stevens and Hall,[5] 0.45, and



0030-3941/78/6801-0140$00.50

# APPENDIX J



# Principles of Optics

## ELECTROMAGNETIC THEORY OF PROPAGATION INTERFERENCE AND DIFFRACTION OF LIGHT

**MAX BORN & EMIL WOLF** **Sixth Edition**

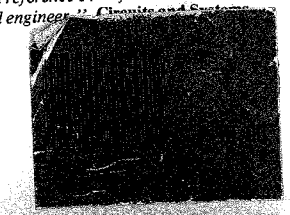
The latest edition of the authors' classic work in the field of Physical Optics incorporating revisions to the text and illustrations as well as references to recent contributions to the literature.

**Contents:** Historical Introduction. Basic Properties of the Electromagnetic Field. Electromagnetic Potentials and Polarization. Foundations of Geometrical Optics. Geometrical Theory of Optical Imaging. Geometrical Theory of Aberrations. Image Forming Instruments. Elements of the Theory of Aberrations. Interference and Diffraction with Partially Coherent Light. Rigorous Diffraction Theory. Diffraction of Light by Ultrasonic Waves. Optics of Metals. Optics of Crystals.

*"Principles of Optics still stands out as the central source of information for any serious student or research worker of optical science."* **Applied Optics**

*"It can be recommended without reserve as the one that every serious worker in optics should have on his bookshelf."* **Optics and Laser**

*"As a result of the rigor, completeness, and elegance in the presentation of classical optics by Born and Wolf, the Principles of Optics will very likely remain for many more years the unsurpassed reference book for the mature, experienced, and serious optical scientist and engineer."* Circuits and S...



BORN & WOLF

Principles of Optics

Sixth Edition

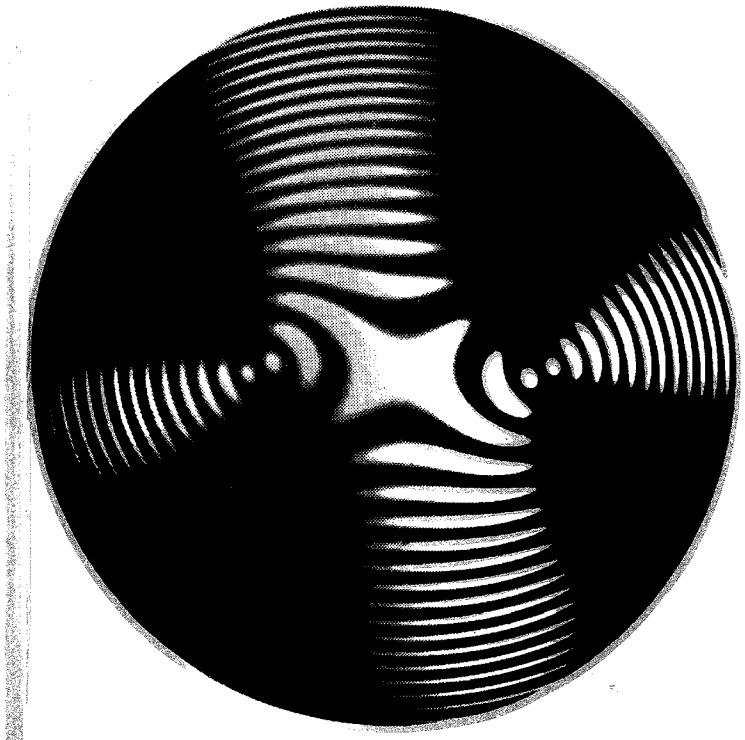
Pergamon

Prin es of Optics

ELECTROMAGNETIC THEORY OF PROPAGATION INTERFERENCE AND DIFFRACTION OF LIGHT

Sixth Edition

N & EMIL WOLF



Pergamon Press

U.K. Pergamon Press plc, Headington Hill Hall, Oxford OX3 0BW, England

U.S.A. Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, NY 10523, U.S.A.

PEOPLE'S REPUBLIC OF CHINA Pergamon Press, Room 4037, Qianmen Hotel, Beijing, People's Republic of China

FEDERAL REPUBLIC OF GERMANY Pergamon Press GmbH, Hammerweg 6, D-6242 Kronberg, Federal Republic of Germany

BRAZIL Pergamon Editora Ltda, Rua Eça de Queiros, 346, CEP 04011, Paraiso, São Paulo, Brazil

AUSTRALIA Pergamon Press Australia Pty Ltd, P.O. Box 544, Potts Point, N.S.W. 2011, Australia

JAPAN Pergamon Press, 5th Floor, Matsuoka Central Building, 1-7-1 Nishishinjuku, Shinjuku-ku, Tokyo 160, Japan

CANADA Pergamon Press Canada Ltd., Suite No. 271, 253 College Street, Toronto, Ontario, Canada M5T 1R5



First edition 1959
Second (revised) edition 1964
Third (revised) edition 1965
Fourth (revised) edition 1970
Fifth (revised) edition 1975
Reprinted 1975, 1977
Sixth edition 1980
Reprinted (with corrections) 1983
Reprinted 1984
Reprinted (with corrections) 1986
Reprinted 1987
Reprinted (with corrections) 1989
Reprinted 1991

**Library of Congress Cataloging in Publication Data**

Born, Max
Principles of optics—6th ed. (with corrections).
1. Optics
I. Title II. Wolf, Emil
535 QC351 80-41470

ISBN 0-08-026482-4 Hardcover
ISBN 0-08-026481-6 Flexicover

*Printed in Great Britain by BPCC Wheatons Ltd, Exeter*

# PREFACE TO THE FIRST EDITION

THE idea of writing this book was a result of frequent enquiries about the possibility of publishing in the English language a book on optics written by one of us* more than twenty-five years ago. A preliminary survey of the literature showed that numerous researches on almost every aspect of optics have been carried out in the intervening years, so that the book no longer gives a comprehensive and balanced picture of the field. In consequence it was felt that a translation was hardly appropriate; instead a substantially new book was prepared, which we are now placing before the reader. In planning this book it soon became apparent that even if only the most important developments which took place since the publication of *Optik* were incorporated, the book would become impracticably large. It was, therefore, deemed necessary to restrict its scope to a narrower field. *Optik* itself did not treat the whole of optics. The optics of moving media, optics of X-rays and $\gamma$ rays, the theory of spectra and the full connection between optics and atomic physics were not discussed; nor did the old book consider the effects of light on our visual sense organ—the eye. These subjects can be treated more appropriately in connection with other fields such as relativity, quantum mechanics, atomic and nuclear physics, and physiology. In this book not only are these subjects excluded, but also the classical molecular optics which was the subject-matter of almost half of the German book. Thus our discussion is restricted to those optical phenomena which may be treated in terms of MAXWELL'S phenomenological theory. This includes all situations in which the atomistic structure of matter plays no decisive part. The connection with atomic physics, quantum mechanics, and physiology is indicated only by short references wherever necessary. The fact that, even after this limitation, the book is much larger than *Optik*, gives some indication about the extent of the researches that have been carried out in classical optics in recent times.

We have aimed at giving, within the framework just outlined, a reasonably complete picture of our present knowledge. We have attempted to present the theory in such a way that practically all the results can be traced back to the basic equations of MAXWELL'S electromagnetic theory, from which our whole consideration starts.

In Chapter I the main properties of the electromagnetic field are discussed and the effect of matter on the propagation of the electromagnetic disturbance is described formally, in terms of the usual material constants. A more physical approach to the question of influence of matter is developed in Chapter II: it is shown that in the presence of an external incident field, each volume element of a material medium may be assumed to give rise to a secondary (scattered) wavelet and that the combination of these wavelets leads to the observable, macroscopic field. This approach is of considerable physical significance and its power is illustrated in a later chapter (Chapter XII) in connection with the diffraction of light by ultrasonic waves, first treated in this way by A. B. BHATIA and W. J. NOBLE; Chapter XII was contributed by Prof. BHATIA himself.

A considerable part of Chapter III is devoted to showing how geometrical optics follows from MAXWELL'S wave theory as a limiting case of short wavelengths. In addition to discussing the main properties of rays and wave-fronts, the vectorial

* MAX BORN, *Optik* (Berlin, Springer, 1933).

For the last zone ($Z_n$) that can be seen from $P$, $QP$ is a tangent to the wave, i.e. $\chi = \pi/2$, and for this value of $\chi$, as already mentioned, $K$ was assumed to be zero Hence $K_n = 0$ and (11) reduces to

$$U(P) = i\lambda K_1 \frac{Ae^{ik(r_0+b)}}{r_0+b} = \tfrac{1}{2}U_1(P), \tag{13}$$

showing that *the total disturbance at $P$ is equal to half of the disturbance due to the first zone.*

Equation (13) is in agreement with the expression for the effect of the spherical wave if

$$i\lambda K_1 = 1,$$

i.e. if

$$K_1 = -\frac{i}{\lambda} = \frac{e^{-i\pi/2}}{\lambda}. \tag{14}$$

The factor $e^{-i\pi/2}$ may be accounted for by assuming that the secondary waves oscillate a quarter of a period out of phase with the primary wave; the other factor can be explained by assuming that the amplitudes of the secondary vibrations are to the amplitudes of the primary vibrations in the ratio $1:\lambda$. We can therefore conclude that, with these assumptions about the amplitude and phase of the secondary waves, the HUYGENS–FRESNEL principle leads to the correct expression for the propagation of a spherical wave in free space. The additional assumptions must, however, be regarded as purely a convenient way of interpreting the mathematical expressions and as being devoid of any physical significance; the real justification of the factor (14) will become evident later (§ 8.3).

Still following FRESNEL, let us consider the effect at $P$ when some of the zones are obstructed by a plane screen with a circular opening, perpendicular to $P_0P$ and with its centre on this line. The total disturbance at $P$ must now be regarded as due to wavelets from only those zones that are not obstructed by the screen. When the screen covers all but half of the first zone, (3) gives, on setting $j = 1$, and multiplying by $\frac{1}{2}$,

$$U(P) = i\lambda K_1 \frac{Ae^{ik(r_0+b)}}{r_0+b} = \frac{Ae^{ik(r_0+b)}}{r_0+b}; \tag{15}$$

hence the disturbance at $P$ is now the same as would be obtained if no screen were present. When all the zones are covered except the first, (3) gives

$$U(P) = 2i\lambda K_1 \frac{Ae^{ik(r_0+b)}}{r_0+b} = 2\frac{Ae^{ik(r_0+b)}}{r_0+b}, \tag{16}$$

so that the intensity $I(P) = |U(P)|^2$ is four times larger than if the screen were absent. When the opening is increased still further the intensity will decrease, since the first two terms in (4) have different signs. Moreover, since $K_1$ and $K_2$ are nearly equal, it follows that there will be almost complete darkness at $P$ when the opening is approximately equal to the first two zones. Thus, when the size of the opening is varied, there is a periodic fluctuation in the intensity at $P$. Similar results are obtained when the size of the opening and the source are fixed, but the position of the point $P$ of observation is varied along the axis; for then, as $P$ gradually approaches the screen, an increasingly larger number of zones is required to fill the opening completely.



All these results were found to be in good agreement with experiment. One prediction of FRESNEL's theory made a strong impression on his contemporaries, and was, in fact, one of the decisive factors which temporarily ended the long battle between the corpuscular and the wave theory of light in favour of the latter. It concerns the effect which arises when the first zone is obstructed by a small circular disc placed at right angles to $P_0P$. According to (5) the complex amplitude at $P$ is then given by

$$U(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b}[-K_2 + K_3 - K_4 + \ldots], \tag{17}$$

and, by a similar argument as before, the sum of the series in the brackets is $-K_2/2$. Since $K_2$ is assumed to differ only slightly from $K_1 = 1/i\lambda$, it follows that there is light in the geometrical shadow of the disc, and, moreover, that the intensity there is the same as if no disc were present.*

## 8.3. KIRCHHOFF'S DIFFRACTION THEORY

### 8.3.1 The integral theorem of Kirchhoff

The basic idea of the HUYGENS–FRESNEL theory is that the light disturbance at a point $P$ arises from the superposition of secondary waves that proceed from a surface situated between this point and the light source. This idea was put on a sounder mathematical basis by KIRCHHOFF†, who showed that the HUYGENS-FRESNEL principle may be regarded as an approximate form of a certain integral theorem‡ which expresses the solution of the homogeneous wave equation, at an arbitrary point in the field, in terms of the values of the solution and its first derivatives at all points on an arbitrary closed surface surrounding $P$.

We consider first a strictly monochromatic scalar wave

$$V(x, y, z, t) = U(x, y, z)e^{-i\omega t}. \tag{1}$$

In vacuum the space-dependent part then satisfies the time-independent wave equation

$$(\nabla^2 + k^2)U = 0, \tag{2}$$

where $k = \omega/c$. Equation (2) is also known as the HELMHOLTZ equation.

* That a bright spot should appear at the centre of the shadow of a small disc was deduced from FRESNEL's theory by S. D. POISSON in 1818. POISSON, who was a member of the committee of the French Academy which reviewed FRESNEL's prize memoir, appears to have considered this conclusion contrary to experiment and so refuting FRESNEL's theory. However, ARAGO, another member of the committee, performed the experiment and found that the surprising prediction was correct. A similar observation had been made a century earlier by MARALDI but had been forgotten.

† G. KIRCHHOFF, *Berl. Ber.* (1882), 641; *Ann. d. Physik.* (2) **18** (1883), 663; *Ges. Abh. Nachtr.*, 29.

KIRCHHOFF's theory applies to the diffraction of scalar waves. As will be shown in § 8.4 a scalar theory is usually quite adequate for the treatment of the majority of problems of instrumental optics.

Vectorial generalizations of the Huygens–Fresnel principle have been proposed by many authors. The first satisfactory generalization is due to F. KOTTLER, *Ann. d. Physik*, **71** (1923), 457; **72** (1923), 320. [Cf. B. B. BAKER and E. T. COPSON, *The Mathematical Theory of Huygens' Principle*, (Oxford, Clarendon Press, 2nd edition, 1950), p. 114].

‡ For monochromatic waves this theorem was derived earlier in acoustics by H. VON HELMHOLTZ, *J. f. Math.*, **57** (1859), 7.

Let $v$ be a volume bounded by a closed surface $S$, and let $P$ be any point within it; we assume that $U$ possesses continuous first- and second-order partial derivatives within and on this surface. If $U'$ is any other function which satisfies the same continuity requirements as $U$, we have by GREEN's theorem

$$\iiint_v (U\nabla^2 U' - U'\nabla^2 U)dv = -\iint_S \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) dS, \tag{3}$$

where $\partial/\partial n$ denotes differentiation along the *inward** normal to $S$. In particular, if $U'$ also satisfies the time-independent wave equation, i.e. if

$$(\nabla^2 + k^2)U' = 0, \tag{4}$$

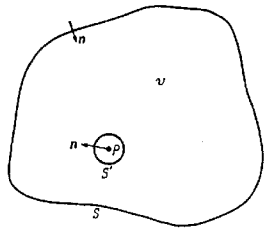



Fig. 8.2. Derivation of the HELMHOLTZ–KIRCHHOFF integral theorem: region of integration.

then it follows at once from (2) and (4) that the integrand on the left of (3) vanishes at every point of $v$, and consequently

$$\iint_S \left(U\frac{\partial U'}{\partial n} - U'\frac{\partial U}{\partial n}\right) dS = 0. \tag{5}$$

Suppose we take $U'(x, y, z) = e^{iks}/s$, where $s$ denotes the distance from $P$ to the point $(x, y, z)$. This function has a singularity for $s = 0$, and since $U'$ was assumed to be continuous and differentiable, $P$ must be excluded from the domain of integration. We shall therefore surround $P$ by a small sphere of radius $\varepsilon$ and extend the integration throughout the volume between $S$ and the surface $S'$ of this sphere (Fig. 8.2). In place of (5), we then have

$$\iint_S + \iint_{S'} \left\{U\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial n}\right\} dS = 0,$$

whence

$$\begin{aligned}\iint_S \left\{U\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial n}\right\} dS &= -\iint_{S'} \left\{U\frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial n}\right\} dS' \\ &= -\iint_\Omega \left\{U\frac{e^{ike}}{\varepsilon}\left(ik - \frac{1}{\varepsilon}\right) - \frac{e^{ike}}{\varepsilon}\frac{\partial U}{\partial s}\right\} \varepsilon^2 d\Omega, \end{aligned} \tag{6}$$

where $d\Omega$ denotes an element of the solid angle. Since the integral over $S$ is independent of $\varepsilon$, we may replace the integral on the right-hand side by its limiting value

* GREEN's theorem is usually expressed in terms of the outward normal, but the inward normal is more convenient in the present application.



as $\varepsilon \to 0$; the first and third terms in this integral give no contribution in the limit, and the total contribution of the second term is $4\pi U(P)$. Hence

$$U(P) = \frac{1}{4\pi}\iint_S \left\{U\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial n}\right\} dS. \tag{7}$$

This is one form of the *integral theorem of Helmholtz and Kirchhoff*.*

We note, that as $k \to 0$, the time-independent wave equation (2) reduces to LAPLACE's equation $\nabla^2 U = 0$, and (7) then goes over into the well-known formula of potential theory

$$U(P) = \frac{1}{4\pi}\iint_S \left\{U\frac{\partial}{\partial n}\left(\frac{1}{s}\right) - \frac{1}{s}\frac{\partial U}{\partial n}\right\} dS. \tag{8}$$

If $P$ lies outside the surface $S$, but $U$ is still assumed to be continuous and differentiable up to the second order within $S$, and if as before we take $U' = e^{iks}/s$, equation (3) remains valid throughout the whole volume within $S$. According to (5) the surface integral then has the value zero.

There is a complementary form of the HELMHOLTZ–KIRCHHOFF theorem for the case when $U$ is continuous and differentiable up to the second order *outside* and on a closed surface $S$ (sources inside). In this case, however, as in other problems of propagation in an infinite medium, the boundary values on $S$ are no longer sufficient to specify the solution uniquely and additional assumptions must be made about the behaviour of the solution as $s \to \infty$. For a discussion of this case we must, however, refer elsewhere.†

So far we have considered strictly monochromatic waves. We now derive the general form of KIRCHHOFF's theorem which applies to waves that are not necessarily monochromatic.

Let $V(x, y, z, t)$ be a solution of the wave equation

$$\nabla^2 V = \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2}, \tag{9}$$

and assume that $V$ can be represented in the form of a FOURIER integral

$$V(x, y, z, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} U_\omega(x, y, z)e^{-i\omega t}\, d\omega. \tag{10}$$

Then, by the FOURIER inversion formula

$$U_\omega(x, y, z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} V(x, y, z, t)e^{i\omega t}\, dt. \tag{11}$$

Since $V(x, y, z, t)$ is assumed to satisfy the wave equation (9), $U_\omega(x, y, z)$ will satisfy the time-independent wave equation (2). If moreover $V$ obeys the appropriate regularity

* This theorem expresses $U(P)$ in terms of the values of both $U$ and $\partial U/\partial n$ on $S$. It may, however, be shown from the theory of GREEN's functions that the values of either $U$ or $\partial U/\partial n$ on $S$ are sufficient to specify $U$ at every point $P$ within $S$. (See for example F. POCKELS: *Über die Partielle Differentialgleichung* $(\nabla^2 + k^2)U = 0$ (Leipzig, Teubner, 1891.) However, only in the simplest cases, e.g. when $S$ is a plane, is it possible to determine the appropriate GREEN's function (cf. A. SOMMERFELD, *Optics* (New York, Academic Press, 1954), p. 199].

† See for example B. B. BAKER and E. T. COPSON, *The Mathematical Theory of Huygens' Principle* (Oxford, Clarendon Press, 2nd ed., 1950), pp. 24–25.

conditions within and on a closed surface $S$, we may apply the KIRCHHOFF formula separately to each FOURIER component $U_\omega(x, y, z) = U_\omega(P)$:

$$U_\omega(P) = \frac{1}{4\pi}\int\int_S \left\{U_\omega \frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U_\omega}{\partial n}\right\} dS. \tag{12}$$

When we change the order of integration and set $k = \omega/c$, (10) becomes,

$$V(P,t) = \frac{1}{4\pi}\int\int_S dS \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left\{U_\omega \frac{\partial}{\partial n}\left(\frac{e^{-i\omega(t-s/c)}}{s}\right) - \frac{e^{-i\omega(t-s/c)}}{s}\frac{\partial U_\omega}{\partial n}\right\} d\omega$$
$$= \frac{1}{4\pi}\int\int_S dS \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left\{U_\omega\left(\frac{\partial}{\partial n}\left(\frac{1}{s}\right) + \frac{i\omega}{sc}\frac{\partial s}{\partial n}\right)e^{-i\omega(t-s/c)} - \frac{e^{-i\omega(t-s/c)}}{s}\frac{\partial U_\omega}{\partial n}\right\} d\omega$$

or using (10),

$$V(P,t) = \frac{1}{4\pi}\int\int_S \left\{[V]\frac{\partial}{\partial n}\left(\frac{1}{s}\right) - \frac{1}{cs}\frac{\partial s}{\partial n}\left[\frac{\partial V}{\partial t}\right] - \frac{1}{s}\left[\frac{\partial V}{\partial n}\right]\right\} dS. \tag{13}$$

The square brackets denote "retarded values", i.e. values of the functions taken at the time $t - s/c$. The formula (13) is the general form of *Kirchhoff's theorem.*

It can also be seen by analogy with the previous case, that the value of the integra in (13) is zero when $P$ is outside $S$.

The last term in (13) represents the contribution of a distribution of sources of strength $-\frac{1}{4\pi}\frac{\partial V}{\partial n}$ per unit area, whilst the first two terms may be shown to represent a contribution of doublets of strength $V/4\pi$ per unit area, directed normally to the surface. Naturally these sources and doublets are fictitious, there being no deep physical significance behind such an interpretation.

### 8.3.2 Kirchhoff's diffraction theory

Whilst the integral theorem of KIRCHHOFF embodies the basic idea of the HUYGENS–FRESNEL principle, the laws governing the contributions from different elements of the surface are more complicated than FRESNEL assumed. KIRCHHOFF showed, however, that in many cases the theorem may be reduced to an approximate but much simpler form, which is essentially equivalent to the formulation of FRESNEL, but which in addition gives an explicit formula for the inclination factor that remained undetermined in FRESNEL's theory.

Consider a monochromatic wave, from a point source $P_0$, propagated through an opening in a plane opaque screen, and let $P$ as before be the point at which the light disturbance is to be determined. We assume that the linear dimensions of the opening, although large compared to the wavelength, are small compared to the distances of both $P_0$ and $P$ from the screen.

To find the disturbance at $P$ we take KIRCHHOFF's integral over a surface $S$ formed by (see Fig. 8.3(a)): (1) the opening $\mathscr{A}$, (2) a portion $\mathscr{B}$ of the non-illuminated side of the screen, and (3) a portion $\mathscr{C}$ of a large sphere of radius $R$, centred at $P$ which, together with $\mathscr{A}$ and $\mathscr{B}$, forms a closed surface.

KIRCHHOFF's theorem, expressed by equation (7), then gives

$$U(P) = \frac{1}{4\pi}\left[\int\int_{\mathscr{A}} + \int\int_{\mathscr{B}} + \int\int_{\mathscr{C}}\right]\left\{U\frac{\partial}{\partial n}\left(\frac{e^{iks}}{s}\right) - \left(\frac{e^{iks}}{s}\right)\frac{\partial U}{\partial n}\right\} dS, \tag{14}$$

where, as before, $s$ is the distance of the element $dS$ from $P$ and $\partial/\partial n$ denotes differentiation along the inward normal to the surface of integration.



The difficulty is encountered that the values of $U$ and $\partial U/\partial n$ on $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$ which should be substituted into (14) are never known exactly. However, it is reasonable to suppose that everywhere on $\mathscr{A}$, except in the immediate neighbourhood of the rim of the opening, $U$ and $\partial U/\partial n$ will not appreciably differ from the values

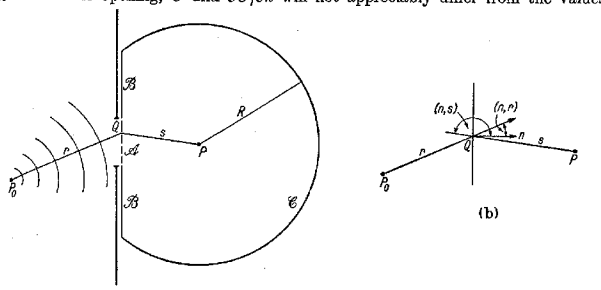



Fig. 8.3. Illustrating the derivation of the FRESNEL–KIRCHHOFF diffraction formula.

obtained in the absence of the screen, and that on $\mathscr{B}$ these quantities will be approximately zero. KIRCHHOFF accordingly set

$$\left.\begin{aligned} &\text{on } \mathscr{A}: \quad U = U^{(i)}, \quad \frac{\partial U}{\partial n} = \frac{\partial U^{(i)}}{\partial n}, \\ &\text{on } \mathscr{B}: \quad U = 0 \qquad\quad \frac{\partial U}{\partial n} = 0, \end{aligned}\right\} \tag{15}$$

where

$$U^{(i)} = \frac{Ae^{ikr}}{r}, \qquad \frac{\partial U^{(i)}}{\partial n} = \frac{Ae^{ikr}}{r}\left[ik - \frac{1}{r}\right]\cos(n, r) \tag{16}$$

are the values relating to the incident field (see Fig. 8.3 (b)) and $A$ is a constant. The approximations (15) are called *Kirchhoff's boundary conditions* and are the basis of *Kirchhoff's diffraction theory.*

It remains to consider the contribution from the spherical portion $\mathscr{C}$. Now it is evident that by taking the radius $R$ sufficiently large, the values of $U$ and $\partial U/\partial n$ on $\mathscr{C}$ may be made arbitrarily small, which suggests that the contribution from $\mathscr{C}$ may be neglected. However, by letting $R$ increase indefinitely, the area of $\mathscr{C}$ also increases beyond all limits, so that the condition $U \to 0$ and $\partial U/\partial n \to 0$ as $R \to \infty$ is not sufficient to make the integral vanish. A more precise assumption about the behaviour of the wave function at a large distance from the screen must therefore be made, a point which we have already touched upon on p. 377 in connection with the uniqueness of solutions in problems involving an infinite medium. For our purposes it is sufficient to make the physically obvious assumption that the radiation field does not exist at all times but that it is produced by a source that begins to radiate at some particular instant of time $t = t_0$.* (This, of course, implies, that we now depart from strict monochromacy, since a perfectly monochromatic field would exist for all

* This assumption is not essential but shortens the discussion. For a more formal argument see M. BORN, *Optik* (Berlin, Springer, 1933), p. 149.

times.) Then at any time $t > t_0$, the field fills a region of space the outer boundary of which is at distance not greater than $c(t - t_0)$ from $P_0$, $c$ being the velocity of light. Hence if the radius $R$ is chosen so large that at the time when the disturbance at $P$ is considered no contributions from $\mathscr{C}$ could have reached $P$ because at the appropriate earlier time the field has not reached these distant regions, the integral over $\mathscr{C}$ will vanish. Thus finally, on substituting into (14), and neglecting in the normal derivatives the terms $1/r$ and $1/s$ in comparison to $k$, we obtain

$$U(P) = -\frac{iA}{2\lambda}\int\int_{\mathscr{A}} \frac{e^{ik(r+s)}}{rs}\,[\cos(n,r) - \cos(n,s)]\,dS, \tag{17}$$

which is known as the *Fresnel–Kirchhoff diffraction formula.*

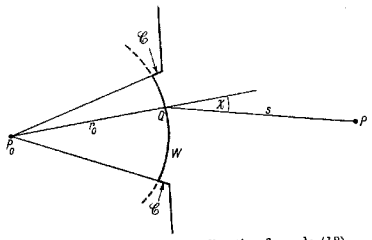



Fig. 8.4. Illustrating the diffraction formula (18).

It is evident that in place of $\mathscr{A}$ any other open surface, the rim of which coincides with the edge of the aperture, could have been chosen. In particular we may choose instead of $\mathscr{A}$ a portion $W$ of an incident wave front which approximately fills the aperture, together with a portion $\mathscr{C}$ of a cone with vertex at $P_0$ and with generators through the rim of the aperture (Fig. 8.4). If the radius of curvature of the wave is sufficiently large, the contribution from $\mathscr{C}$ may obviously be neglected. Also, on $W$, $\cos(n, r_0) = 1$. If further we set $\chi = \pi - (r_0, s)$ we obtain, in place of (17),

$$U(P) = -\frac{i}{2\lambda}\frac{Ae^{ikr_0}}{r_0}\int\int_W \frac{e^{iks}}{s}\,(1 + \cos\chi)\,dS, \tag{18}$$

where $r_0$ is the radius of the wave-front $W$. This result is in agreement with FRESNEL's formulation of HUYGENS' principle if, as the contribution from the element $dW$ of the wave-front we take

$$-\frac{i}{2\lambda}\frac{Ae^{ikr_0}}{r_0}\frac{e^{iks}}{s}\,(1 + \cos\chi)\,dS. \tag{19}$$

Comparison of (18) with § 8.2 (1) gives for the inclination factor of FRESNEL's theory the expression*

$$K(\chi) = -\frac{i}{2\lambda}\,(1 + \cos\chi) \tag{20}$$

For the central zone $\chi = 0$, and (20) gives $K_1 = K(0) = -i/\lambda$ in agreement with § 8.2. (14). It is however not true, as FRESNEL assumed, that $K(\pi/2) = 0$.

* Expression (20) for the inclination factor was first derived by G. G. STOKES, *Trans. Camb. Phil. Soc.*, **9** (1849), 1; reprinted in his *Math. and Phys. Papers* (Cambridge University Press, 1883), 243).



Returning now to the FRESNEL–KIRCHHOFF diffraction formula (17), we note that it is symmetrical with respect to the source and the point of observation. This implies that *a point source at $P_0$ will produce at $P$ the same effect as a point source of equal intensity placed at $P$ will produce at $P_0$.* This result is sometimes referred to as the *reciprocity theorem* (or the *reversion theorem*) *of Helmholtz.*

So far we have assumed that the light on its passage from the source to $P$ does not encounter any other surface than the diffracting screen; the incident waves are then spherical. The analysis can be easily extended to cover more complicated cases, where the waves are no longer of such simple form. It is again found that, provided the radii of curvature at each point of the wave-front are large compared to the wave-length, and provided that the angles involved are sufficiently small, the results of KIRCHHOFF's theory are substantially equivalent to predictions based on the HUYGENS–FRESNEL principle.

From the preceding discussion we can also immediately draw a conclusion which concerns the distribution of light diffracted by complementary screens, i.e. by screens which are such that the openings in one correspond exactly to the opaque portions of the others and vice versa. Let $U_1(P)$ and $U_2(P)$ denote respectively the values of the complex displacement when the first or the second screen alone is placed between the source and the point $P$ of observation, and let $U(P)$ be the value when no screen is present. Then, since $U_1$ and $U_2$ can be expressed as integrals over the openings, and since the openings in the two screens just add up to fill the whole plane,

$$U_1 + U_2 = U. \tag{21}$$

This result is known as *Babinet's principle.**

From BABINET's principle two conclusions follow at once: If $U_1 = 0$, then $U_2 = U$; hence at points at which the intensity is zero in the presence of one of the screens, the intensity in the presence of the other is the same as if no screen was present. Further if $U = 0$, then $U_1 = -U_2$; this implies that, at points where $U$ is zero, the phases of $U_1$ and $U_2$ differ by $\pi$ and the intensities $I_1 = |U_1|^2$, $I_2 = |U_2|^2$ are equal. If for example a point source is imaged by an error-free lens, the light distribution $U$ in the image plane will be zero except in the immediate neighbourhood of the image $O$ of the source. If then complementary screens are placed between the object and the image one has $I_1 = I_2$ except in the neighbourhood of $O$.

The consequences of the basic approximation (15) of KIRCHHOFF's theory have been subject to many critical discussions, which showed, for example, that KIRCHHOFF's solution does not reproduce the assumed values in the plane of the aperture†. However, more recently it was shown by WOLF and MARCHAND‡ that KIRCHHOFF's theory

* A. BABINET, *Compt. Rend.*, **4** (1837), 638. An analogous theorem of this type, which involves the electromagnetic field vectors rather than the single scalar $U$ and which may be considered as rigorous formulation of BABINET's principle, is given in § 11.3.

† H. POINCARÉ, *Théorie mathématique de la lumière* (Paris, George Carré, II (1892)), pp. 187–8. See also B. B. BAKER and E. T. COPSON, *The Mathematical Theory of Huygens' Principle* (Oxford, Clarendon Press, 2nd ed., 1950), pp. 71–72 and G. TORALDO DI FRANCIA, *Atti Fond. Giorgio Ronchi*, XI (1956), § 6.

‡ E. WOLF and E. W. MARCHAND, *J. Opt. Soc. Amer.*, **56** (1966), 1712. Also, it has been shown by F. KOTTLER, *Ann. der Physik*, **70** (1923), 405 that KIRCHHOFF's theory may be regarded as providing a rigorous solution to a certain saltus problem (problem with prescribed discontinuities rather than prescribed boundary values). This interpretation is of particular interest in connection with the problem of diffraction at a black (completely absorbing) screen. See also F. KOTTLER, *Progress in Optics*, Vol. 4, ed. E. WOLF (Amsterdam, North Holland Publishing Company and New York, J. Wiley and Sons, 1964), p. 281 and B. B. BAKER and E. T. COPSON, *loc. cit.*, p. 98.]

An article by C. J. BOUWKAMP, *Rep. Progr. Phys.* (London, Physical Society), **17** (1954), 35, contains references to numerous papers concerned with various modifications of KIRCHHOFF's theory.

may be interpreted in a mathematically consistent way, as providing an exact solution to a somewhat different boundary value problem than that specified by eqs. (15) and (16). It turns out that KIRCHHOFF's theory is entirely adequate for the treatment of the majority of problems encountered in instrumental optics. This is mainly due to the smallness of the optical wavelengths in comparison with the dimensions of the diffracting obstacles.* In other problems, such as those relating to the behaviour of the field in the immediate neighbourhood of screens and obstacles, more refined methods have to be used; they must then be considered as boundary-value problems of electromagnetic theory, with the sources as appropriate singularities of the wave functions. Only in a very limited number of cases have such solutions been found; some of them will be discussed in Chapter XI.

### 8.3.3 Fraunhofer and Fresnel diffraction

We now examine more closely the FRESNEL–KIRCHHOFF diffraction integral (17),

$$U(P) = -\frac{Ai}{2\lambda}\iint_{\mathscr{A}} \frac{e^{ik(r+s)}}{rs}[\cos(n,r) - \cos(n,s)]\,dS. \tag{22}$$

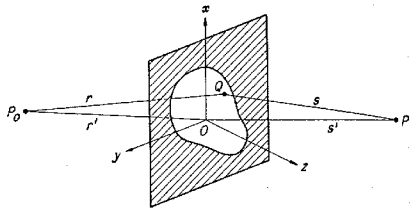



Fig. 8.5. Diffraction at an aperture in a plane screen.

As the element $dS$ explores the domain of integration, $r + s$ will in general change by very many wavelengths, so that, the factor $e^{ik(r+s)}$ will oscillate rapidly. On the other hand, if the distances of the points $P_0$ and $P$ from the screen are large compared to the linear dimensions of the aperture, the factor $[\cos(n, r) - \cos(n, s)]$ will not vary appreciably over the aperture. Further, we assume that if $O$ is any point in the aperture, the angles which the lines $P_0O$ and $OP$ make with $P_0P$ are not too large. We may then replace this factor by $2\cos\delta$, where $\delta$ is the angle between the line $P_0P$ and the normal to the screen. Finally the factor $1/rs$ may be replaced by $1/r's'$, where $r'$ and $s'$ are the distance of $P_0$ and $P$ from the origin and (22) then reduces to

$$U(P) \sim -\frac{Ai}{\lambda}\frac{\cos\delta}{r's'}\iint_{\mathscr{A}} e^{ik(r+s)}\,dS. \tag{23}$$

We take a Cartesian reference system with origin in the aperture and with the $x$— and $y$ — axes in the plane of the aperture and choose the positive $z$ direction to point into the half-space that contains the point $P$ of observation (Fig. 8.5).

If $(x_0, y_0, z_0)$ and $(x, y, z)$ are the coordinates of $P_0$ and of $P$ respectively, and $(\xi, \eta)$ the coordinates of a point $Q$ in the aperture, we have

$$\left.\begin{aligned} r^2 &= (x_0 - \xi)^2 + (y_0 - \eta)^2 + z_0^2, \\ s^2 &= (x - \xi)^2 + (y - \eta)^2 + z^2, \end{aligned}\right\} \tag{24}$$

* Cf. S. SILVER, *J. Opt. Soc. Amer.*, **52** (1962), 131.



$$\left.\begin{aligned} r'^2 &= x_0^2 + y_0^2 + z_0^2, \\ s'^2 &= x^2 + y^2 + z^2. \end{aligned}\right\} \tag{25}$$

Hence

$$\left.\begin{aligned} r^2 &= r'^2 - 2(x_0\xi + y_0\eta) + \xi^2 + \eta^2, \\ s^2 &= s'^2 - 2(x\xi + y\eta) + \xi^2 + \eta^2. \end{aligned}\right\} \tag{26}$$

Since we assumed that the linear dimensions of the aperture are small compared to both $r'$ and $s'$ we may expand $r$ and $s$ as power series in $\xi/r'$, $\eta/r'$, $\xi/s'$ and $\eta/s'$. We then obtain

$$\left.\begin{aligned} r &\sim r' - \frac{x_0\xi + y_0\eta}{r'} + \frac{\xi^2 + \eta^2}{2r'} - \frac{(x_0\xi + y_0\eta)^2}{2r'^3} - \ldots \\ s &\sim s' - \frac{x\xi + y\eta}{s'} + \frac{\xi^2 + \eta^2}{2s'} - \frac{(x\xi + y\eta)^2}{2s'^3} - \ldots \end{aligned}\right\} \tag{27}$$

Substitution from (27) into (23) gives

$$U(P) = -\frac{i\cos\delta}{\lambda}\frac{Ae^{ik(r'+s')}}{r's'}\iint_{\mathscr{A}} e^{ikf(\xi,\eta)}\,d\xi d\eta, \tag{28}$$

where

$$f(\xi,\eta) = -\frac{x_0\xi + y_0\eta}{r'} - \frac{x\xi + y\eta}{s'} + \frac{\xi^2 + \eta^2}{2r'} + \frac{\xi^2 + \eta^2}{2s'} - \frac{(x_0\xi + y_0\eta)^2}{2r'^3} - \frac{(x\xi + y\eta)^2}{2s'^3} \ldots \tag{29}$$

If we denote by $(l_0, m_0)$ and $(l, m)$ the first two direction cosines

$$\left.\begin{aligned} l_0 &= -\frac{x_0}{r'}, \qquad l = \frac{x}{s'}, \\ m_0 &= -\frac{y_0}{r'}, \qquad m = \frac{y}{s'}, \end{aligned}\right\} \tag{30}$$

(29) may be written in the form

$$f(\xi,\eta) = (l_0 - l)\xi + (m_0 - m)\eta + \tfrac{1}{2}\left\{\left(\frac{1}{r'} + \frac{1}{s'}\right)(\xi^2 + \eta^2) - \frac{(l_0\xi + m_0\eta)^2}{r'} - \frac{(l\xi + m\eta)^2}{s'}\right\} \ldots \tag{31}$$

We have reduced the problem of determining the light disturbance at $P$ to the evaluation of the integral (28). Naturally the evaluation is simpler to carry out when the quadratic and higher order terms in $\xi$ and $\eta$ may be neglected in $f$. In this case one speaks of *Fraunhofer diffraction*; when the quadratic terms cannot be neglected, one speaks of *Fresnel diffraction*. Fortunately the simpler case of FRAUNHOFER diffraction is of much greater importance in optics.

Strictly speaking, the second and higher order terms disappear only in the limiting case $r' \to \infty$, $s' \to \infty$, i.e. when both the source and the point of observation are at infinity; (the factor $A$ outside the integral must then be assumed to tend to infinity

# APPENDIX K

Reprinted from *Wavefront Sensing*, Proc. SPIE Vol. 351, pp. 42-53 (1982).


# Comparison of wavefront sensor configurations using optimal reconstruction and correction

**Edward P. Wallner**
Itek Optical Systems, a Division of Itek Corporation
10 Maguire Road, Lexington, Massachusetts 02173

## Abstract

Recent work on optimal reconstruction of wavefronts from slope measurements has developed methods based on more realistic models of sensor devices. These methods also resolve certain ambiguities present in earlier reconstruction methods. In the present paper, these methods are applied to sensor configurations which have been proposed in the literature and the errors in estimating and correcting wavefronts are compared.

## Current wavefront reconstruction methods

Most means of defining the shape of a wavefront are based on measurements of local wavefront slope. For instance, in a shearing interferometer, two displaced copies of the wavefront are interfered and the phase of the interference pattern is a measure of wavefront slope in the direction of the displacement or shear. In a Hartmann test, the displacement of a spot from its nominal position is also a measure of local slope.

The wavefront itself must be reconstructed from the slope data to evaluate or correct its errors. One scheme for this was described by Rimmer.[1] Here the measurements are made by arrays of x and y slope sensors which are displaced as shown in Figure 1. The output of each sensor then approximates the phase difference between the grid points on the edges of the sensor. This ties the phases at the grid points together in a network such as that in Figure 2, which corresponds to the configuration of Figure 1. There are many redundant paths in this network allowing the measurement noise to be reduced by averaging.

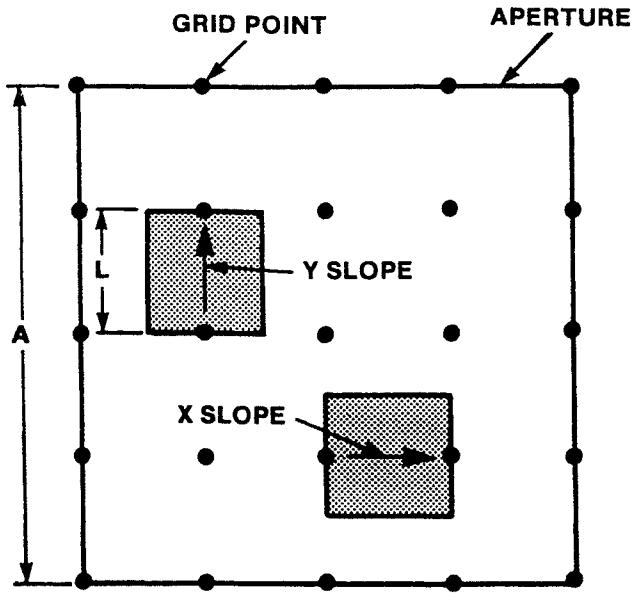



Figure 1. Displaced subaperture sensor configuration.

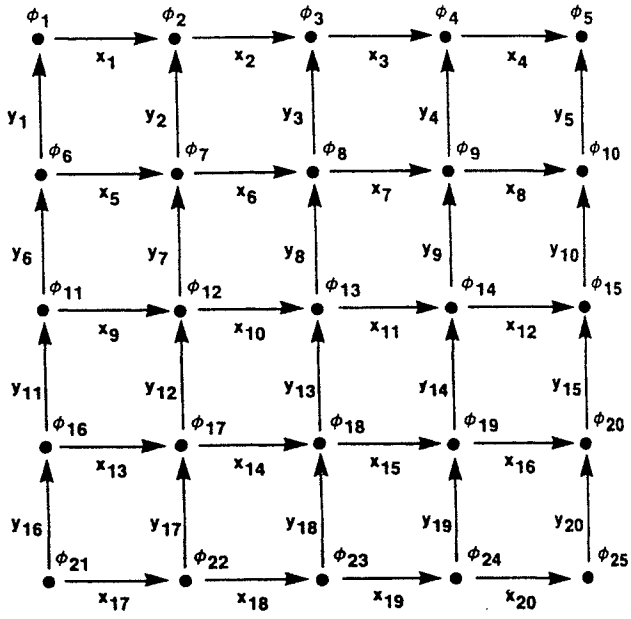



Figure 2. Network for phase reconstruction with displaced subapertures.

If only measurement noise is considered and it is equal and independent in each sensor, the phase estimate giving the least mean square error is the average of the estimates obtained from the phase at adjacent points plus the measured phase differences.[1,2,3]

$$\phi_j = \frac{1}{N} \sum_i \phi_i + \Delta\phi_{ij} \qquad (1)$$

where $\phi_j$ = phase at j-th grid point

$\Delta\phi_{ij}$ = measured phase difference between grid points i and j

The solution to this system of equations is generally found by iteration, and it may be expressed as a matrix that multiplies the vector of phase differences to produce the vector of phases

$$\phi_j = M_{jk}\ \Delta\phi_k \qquad (2)$$

where $M_{jk}$ = solution matrix

$\Delta\phi_k$ = k-th phase difference measurement

The average phase over the aperture is set to zero, since a constant phase will neither affect optical performance nor be detected by the slope sensor.

A different wavefront sensor configuration, representing a Hartmann type of sensor, was analyzed by Fried.[4] This sensor, as shown in Figure 3, measures the x and y slopes in the same subaperture. If the measurements are resolved onto axes rotated by 45° (or if the measurements are made in those coordinates), the new measurements approximate the phase differences between the corners of the subapertures. Figure 4 shows the resulting network for the configuration in Figure 3. There are now two interlaced grids of points that share no common grid point. The phase can be reconstructed for each net as before, but each has an arbitrary constant phase term. The overall average phase can again be set to zero but the difference between the two nets remains as a degree of freedom.[5] A difference in the average phases will lead to a "checker board" phase pattern to which the sensor is insensitive.

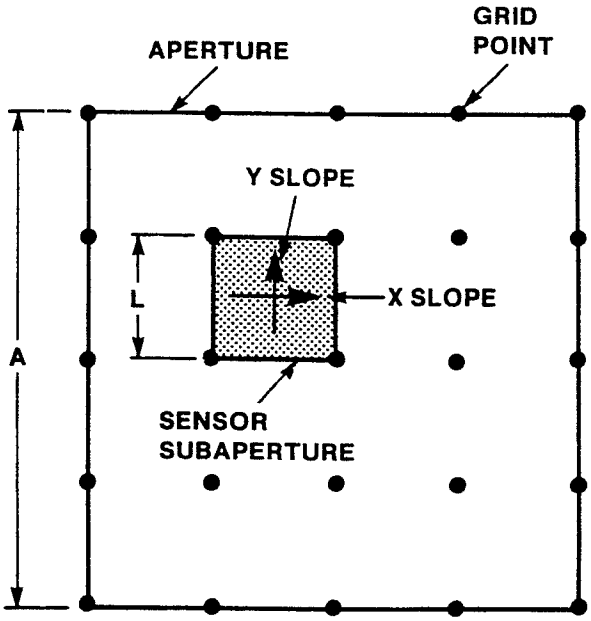



Figure 3. Common subaperture sensor configuration.

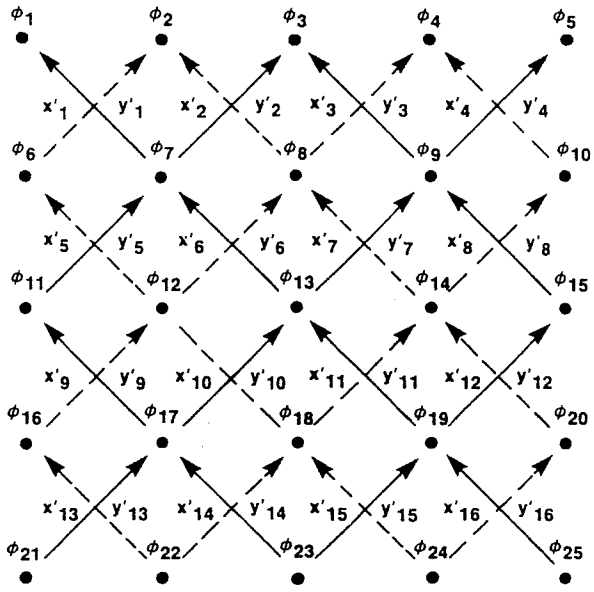



Figure 4. Network for phase reconstruction with common subapertures.

In addition to the arbitrary degree of freedom in the latter case, there are several other limitations to this approach to wavefront reconstruction:

1. The slope measurement is treated as a point-to-point phase difference rather than an average over a finite area of the aperture;
2. Estimates are obtained for the grid points only with no guidance on how to interpolate between them; and
3. No use is made of the statistics of the wavefront being estimated.

In the next section, a reconstruction method that overcomes these limitations is developed.

## Optimal wavefront reconstruction from slope measurements

The system considered here consists of an aperture and a number of wavefront sensors, each of which measures the weighted average of wavefront slope over some region within the aperture. The aim is to use the measurements produced by these sensors to construct an estimate of wavefront phase throughout the aperture with minimum mean squared error.

The aperture of the optical system can be described by a weighting function proportional to the intensity of light at the wavefront. It is convenient to normalize this function so that

$$\int_{-\infty}^{\infty} d\bar{x}\ W_A(\bar{x}) = 1 \tag{3}$$

where $\bar{x}$ = two-dimensional vector position in the aperture plane

$W_A(\bar{x})$ = aperture weighting function

$\int_{-\infty}^{\infty} d\bar{x}$ = indicates integration over the entire aperture plane

The wavefront phase or optical path difference is described by the function $\psi(\bar{x})$. Because any uniform phase over the aperture does not affect optical performance, we can equivalently deal with the phase with aperture average removed, $\phi(\bar{x})$. For this function

$$\int_{-\infty}^{\infty} d\bar{x}\ W_A(\bar{x})\ \phi(\bar{x}) = 0 \tag{4}$$

where $\phi(\bar{x}) = \psi(\bar{x}) - \int_{-\infty}^{\infty} d\bar{x}'\ W_A(\bar{x}^1)\ \Psi(\bar{x}')$

$\psi(\bar{x})$ = input wavefront phase

Each wavefront sensor forms a weighted average of the wavefront slope over some region of the aperture. This average is corrupted by additive noise. The signal out of the n-th sensor is then

$$s_n = \int_{-\infty}^{\infty} d\bar{x}\ W_{sn}(\bar{x})\ [\phi_n'(\bar{x}) + v(\bar{x})] \tag{5}$$

where $s_n$ = signal from n-th sensor

$W_{sn}(\bar{x})$ = weighting function for n-th sensor

$\phi_n'(x)$ = slope of wavefront in the direction of sensitivity of n-th sensor

$v(\bar{x})$ = noise distribution over aperture

The signal can be expressed in terms of the wavefront phase by integrating Eq. (5) by parts:

$$s_n = \int_{-\infty}^{\infty} d\bar{x}\ [-W_{sn}'(\bar{x})\ \phi(\bar{x}) + W_{sn}(\bar{x})\ v(\bar{x})] \tag{6}$$

In Eq. (6) use has been made of the fact that $W_{sn}(\bar{x})$ goes to zero at infinity. The noise is assumed to be unbiased, and uncorrelated with wavefront phase

$$\langle v(\bar{x}) \rangle = \langle v(\bar{x})\ \phi(\bar{x}') \rangle = 0 \tag{7}$$

The estimate of phase at some point $\bar{x}$ in the aperture is taken as a linear combination of the slope measurements

$$\hat{\phi}(\bar{x}) = \sum_n s_n\ r_n(\bar{x}) \tag{8}$$

where $\hat{\phi}(\bar{x})$ = estimated phase at $\bar{x}$

$r_n(\bar{x})$ = n-th weighting function in estimate

We wish to determine a set of weighting functions $r_n(\bar{x})$ that minimize the error in the estimate. The error is

$$\begin{aligned} \varepsilon(\bar{x}) &= \hat{\phi}(\bar{x}) - \phi(\bar{x}) \\ &= \sum_n s_n\ r_n(\bar{x}) - \phi(\bar{x}) \end{aligned} \tag{9}$$

The mean square error in the estimate is then

$$\begin{aligned} \langle \varepsilon^2(\bar{x}) \rangle = &\sum_n \sum_{n'} r_n(\bar{x})\ r_{n'}(\bar{x})\ \langle s_n s_{n'} \rangle \\ &- \sum_n r_n(\bar{x})\ \langle s_n \phi(\bar{x}) \rangle \\ &- \sum_{n'} r_{n'}(\bar{x})\ \langle s_{n'} \phi(\bar{x}) \rangle + \langle \phi^2(\bar{x}) \rangle \end{aligned} \tag{10}$$

where $\langle f \rangle$ = indicates the expected value of f

Eq. (10) can be simplified by defining

$$\begin{aligned} S_{nn'} &= \langle s_n s_{n'} \rangle \\ &= \int_{-\infty}^{\infty} d\bar{x}' \int_{-\infty}^{\infty} d\bar{x}'' \left[ W'_{sn}(\bar{x}')\ W'_{sn'}(\bar{x}'')\ \langle \phi(\bar{x}')\ \phi(\bar{x}'') \rangle \right. \\ &\quad \left. + W_{sn}(\bar{x}')\ W_{sn'}(\bar{x}'')\ \langle v(\bar{x}')\ v(\bar{x}'') \rangle \right] \end{aligned} \tag{11}$$

$$\begin{aligned} f_n(\bar{x}) &= \langle s_n \phi(\bar{x}) \rangle \\ &= \int_{-\infty}^{\infty} d\bar{x}'\ W'_{sn}(\bar{x}')\ \langle \phi(\bar{x})\ \phi(\bar{x}') \rangle \end{aligned} \tag{12}$$

Eq. (7) has been used in reducing Eqs. (11) and (12).

Substituting into Eq. (10) gives the simplified expression for the mean square error at $\bar{x}$:

$$\begin{aligned} \langle \varepsilon^2(\bar{x}) \rangle = &\sum_n \sum_{n'} r_n(\bar{x})\ r_{n'}(x)\ S_{nn'} \\ &- \sum_n r_n(\bar{x})\ f_n(\bar{x}) \\ &- \sum_{n'} r_{n'}(\bar{x})\ f_{n'}(\bar{x}) + \langle \phi^2(\bar{x}) \rangle \end{aligned} \tag{13}$$

The optimum value of $r_n(x)$ is found by differentiating Eq. (13) with respect to $r_{n'}(x)$ and equating to zero:

$$d\,\langle\varepsilon^2(\bar{x})\rangle/dr_{n'}(\bar{x}) = \sum_n r_n(\bar{x})\; S_{nn'} - f_{n'}(\bar{x})$$

$$= 0 \tag{14}$$

The solution for the optimum estimation function is found by inverting Eq. (14). Indicating the optimum by a superscript * yields

$$r_n^*(\bar{x}) = \sum_{n'} S_{nn'}^{-1}\, f_{n'}(\bar{x}) \tag{15}$$

where $r_n^*(x)$ = optimum estimation function for n-th sensor

$S_{nn'}^{-1}$ = inverse of $S_{nn'}$ matrix

The matrix $S_{nn'}$ will not be singular unless there are redundant slope measurements with the same noise, which would not occur in any real system. However, as the number of sensors is increased and the noise is reduced, the matrix becomes less well conditioned and special inversion techniques may be required to maintain accuracy in computation.

The resulting minimum mean square error at $\bar{x}$ is found by substituting in Eq. (13):

$$\langle\varepsilon^2(\bar{x})\rangle^* = \langle\phi^2(\bar{x})\rangle - \sum_n r_n^*(\bar{x})\; f(\bar{x}) \tag{16}$$

Finally, the minimum mean square error averaged over the aperture is

$$\sigma_R^2 = \int_{-\infty}^{\infty} d\bar{x}\; W_A(\bar{x})\; \langle\varepsilon^2(\bar{x})\rangle^* \tag{17}$$

The accuracy of the estimate will depend on the signal-to-noise ratio of the measurements. With very noisy data, the matrix $S_{nn'}$ becomes large and the estimated wavefront becomes small as the a priori estimate of zero phase becomes more reliable than the measurements. In this limit, the error approaches the original wavefront and is independent of the density of sensors.

For small measurement errors the matrix $S_{nn'}$ approaches a fixed value and the estimation error approaches a limit set by the inability of the finite array of sensors to measure or infer the wavefront perfectly.

Representation by structure function

In the case of atmospheric turbulence, the statistics of the input wavefront are defined in terms of a structure function

$$D(\bar{x}, \bar{x}') = \langle[\psi(\bar{x}) - \psi(\bar{x}')]^2\rangle \tag{18}$$

The correlation of phases with averages removed required in the computation of $\langle\phi^2(\bar{x})\rangle$, $S_{nn'}$, and $f_n(\bar{x})$ can be derived directly using Eq. (18). The average removed phase is

$$\phi(\bar{x}) = \int_{-\infty}^{\infty} d\bar{x}''\; W_A(\bar{x}'')\; [\psi(\bar{x}) - \psi(\bar{x}'')] \tag{19}$$

The correlation required is

$$\langle \phi(\bar{x})\ \phi(\bar{x}') \rangle$$

$$= \int_{-\infty}^{\infty} d\bar{x}'' \int_{-\infty}^{\infty} d\bar{x}''' \ W_A(\bar{x}'')\ W_A(\bar{x}''') \ \langle [\psi(\bar{x}) - \psi(\bar{x}'')]\ [\psi(\bar{x}') - \psi(\bar{x}''')] \rangle$$

$$= -\tfrac{1}{2} \int_{-\infty}^{\infty} d\bar{x}'' \int_{-\infty}^{\infty} d\bar{x}''' \ W_A(\bar{x}'')\ W_A(\bar{x}''')$$

$$[D(\bar{x}, \bar{x}') - D(\bar{x}, \bar{x}''') - D(\bar{x}'', \bar{x}') + D(\bar{x}'', \bar{x}''')]$$

$$= -\tfrac{1}{2} D(\bar{x}, \bar{x}') + g(\bar{x}) + g(\bar{x}') - a, \tag{20}$$

where $$g(\bar{x}) = \tfrac{1}{2} \int_{-\infty}^{\infty} d\bar{x}'' \ W_A(\bar{x}'')\ D(\bar{x}, \bar{x}'') \tag{21}$$

$$a = \tfrac{1}{2} \int_{-\infty}^{\infty} d\bar{x}'' \int_{-\infty}^{\infty} d\bar{x}''' \ W_A(\bar{x}'')\ W_A(\bar{x}''')\ D(\bar{x}'', \bar{x}''')$$

$$= \int_{-\infty}^{\infty} d\bar{x}\ W_A(\bar{x})\ g(\bar{x}) \tag{22}$$

The mean square wavefront phase at $\bar{x}$ is

$$\langle \phi^2(\bar{x}) \rangle = 2g(\bar{x}) - a \tag{23}$$

The aperture average of the mean square phase is

$$\int_{-\infty}^{\infty} d\bar{x}\ W_A(\bar{x})\ \langle \phi^2(\bar{x}) \rangle = a \tag{24}$$

In the evaluation of $S_{nn'}$ by substituting in Eq. (11) all terms on the right in Eq. (20) except the first can be seen to integrate to zero, leaving

$$S_{nn'} = \int_{-\infty}^{\infty} d\bar{x}' \int_{-\infty}^{\infty} d\bar{x}'' \ [-\tfrac{1}{2}\ W'_{sn}(\bar{x}')\, W'_{sn}(\bar{x}'')\ D(\bar{x}', \bar{x}'')$$

$$+ W_{sn}(\bar{x}')\ W_{sn'}(\bar{x}'')\ \langle v(\bar{x}')\ v(\bar{x}'') \rangle] \tag{25}$$

The function $f_n(\bar{x})$ is evaluated by substituting Eq. (20) in Eq. (12).

$$f_n(\bar{x}) = \int_{-\infty}^{\infty} d\bar{x}'\ W'_{sn}(\bar{x}')\, [\tfrac{1}{2} D(\bar{x}, \bar{x}') - g(\bar{x}')]$$

$$= \tfrac{1}{2} \int_{-\infty}^{\infty} d\bar{x}'\ W'_{sn}(\bar{x}')\ D(\bar{x}, \bar{x}') - c_n \tag{26}$$

The constant $c_n$ can be seen to be the value of the first term on the right averaged over the aperture, so that the function $f_n(\bar{x})$ has no net average phase. The same will also be true of the estimation functions $r_n^*(\bar{x})$.

Photon noise

In many systems of interest, the limit on performance is set by the noise due to photon statistics in the wavefront measurements. The photon noise for a single wavefront can be represented by making $v(\bar{x})$ a spatially white noise function. The correlation of the noise on two wavefront measurements depends in detail on the configuration of the wavefront sensor used. If the two measurements are made on separate wavefronts, either in sequence or separated by a beam splitter, they will be uncorrelated even though they refer to the same area in the aperture.

Measurements of orthogonal slopes made simultaneously on the same wavefront will also be uncorrelated. To reflect these conditions, the correlation of the noise functions may be written

$$\langle v(\bar{x})\ v(\bar{x}')\rangle = k_{nn'}\,\sigma_n^2\,\delta(\bar{x}-\bar{x}') \tag{27}$$

where $\sigma_n^2$ = photon noise density

$k_{nn'} = \bar{l}_n \cdot \bar{l}_{n'}$, for n and n′ measured simultaneously on the same area of the same wavefront

= 0 otherwise

$\bar{l}_n$ = unit vector in direction of slope sensitivity of n-th measurement

$\delta(x)$ = Dirac delta function

The noise term in Eqs. (11) or (25) may then be written

$$\int_{-\infty}^{\infty} d\bar{x}' \int_{-\infty}^{\infty} d\bar{x}'' \, W_{sn}(\bar{x}')\ W_{sn}(\bar{x}'')\ \langle v(\bar{x}')\ v(\bar{x}'')\rangle$$

$$= k_{nn'}\,\sigma_n^2 \int_{-\infty}^{\infty} dx'\ W_{sn}(\bar{x}')\ W_{sn'}(\bar{x}') \tag{28}$$

In most cases the basic measurements will be disjoint and $k_{nn'}$ would be the Kronecker delta if each measurement is treated separately. It may be convenient to combine a number of the basic measurements before processing in order to reduce the order of the matrices involved, and in this case the more general form of $k_{nn'}$ would be needed.

## Optimal wavefront correction

The analysis above applies directly to the problem of how best to correct a wavefront given the outputs of a set of wavefront sensors. If each sensor drives an actuator that has a response of $r_n^*(\bar{x})$, the total correction of $\hat{\phi} = \sum_n s_n r_n^*(\bar{x})$ will give the best possible correction for a linear system. No additional degrees of freedom beyond the number of sensors can improve the correction.

The commands to the correction functions developed here are not in general statistically independent so that if any corrections are omitted, those remaining would not give an optimal correction with the reduced number of degrees of freedom. By properly transforming the commands, however, they can be made independent, allowing the best correction for any number of degrees of freedom to be found.

If $G_{nn'}$ is an orthogonal matrix consisting of unit eigenvectors of the $S_{nn}^{-1}$ matrix, the transformed commands will be

$$s_n' = \sum_{n'} G_{nn'}\, s_{n'} \tag{29}$$

and then optimum transformed correction functions will be

$$r_n^{*\prime}(x) = \sum_{n'} G_{nn'}\, r_{n'}^*(\bar{x}) \tag{30}$$

The reduction in the mean square residual error averaged over the aperture attributable to the n-th transformed correction is

$$\Delta_n\sigma_R^2 = \int_{-\infty}^{\infty} d\bar{x}\ W_A(\bar{x})\ f_n'(\bar{x})\ r_n^{*\prime}(\bar{x}) \tag{31}$$

where

$$f_n'(\bar{x}) = \sum_{n'} G_{nn'}\, f_{n'}(\bar{x}) \tag{32}$$

By adding correction functions in the order of magnitude of their overall error reduction, the minimum expected residual error is obtained for any number of degrees of freedom.

With low noise and well distributed sensors, the lower order correction functions will be similar to those of the Karhunen-Loeve expansion that corresponds to the case of continuous, noise-free measurements of the wavefront. This expansion has been developed for a circular aperture and a Kolmogorov structure function by Wang and Markey.[6]

The case of the best possible correction given a specified set of non-optimal correctors has been treated elsewhere.[7]

## Computational results

The theory derived here has been applied to a unit square aperture with uniform illumination with wavefront sensors that measure the average slope with uniform weight over square subapertures filling the aperture. The measurement noise was taken to be photon noise.

Three configurations of sensors were considered corresponding to the three cases treated by Southwell.[8] These are shown in Figure 5 for the case in which the subaperture dimension L is half the aperture dimension A. In the first configuration, as shown in Figures 1 and 5a, x and y slopes were measured in displaced apertures. The sensors at an edge parallel to the direction of the slope measurement have half of their area outside the aperture, and the signal and photon noise power are accordingly halved for them.

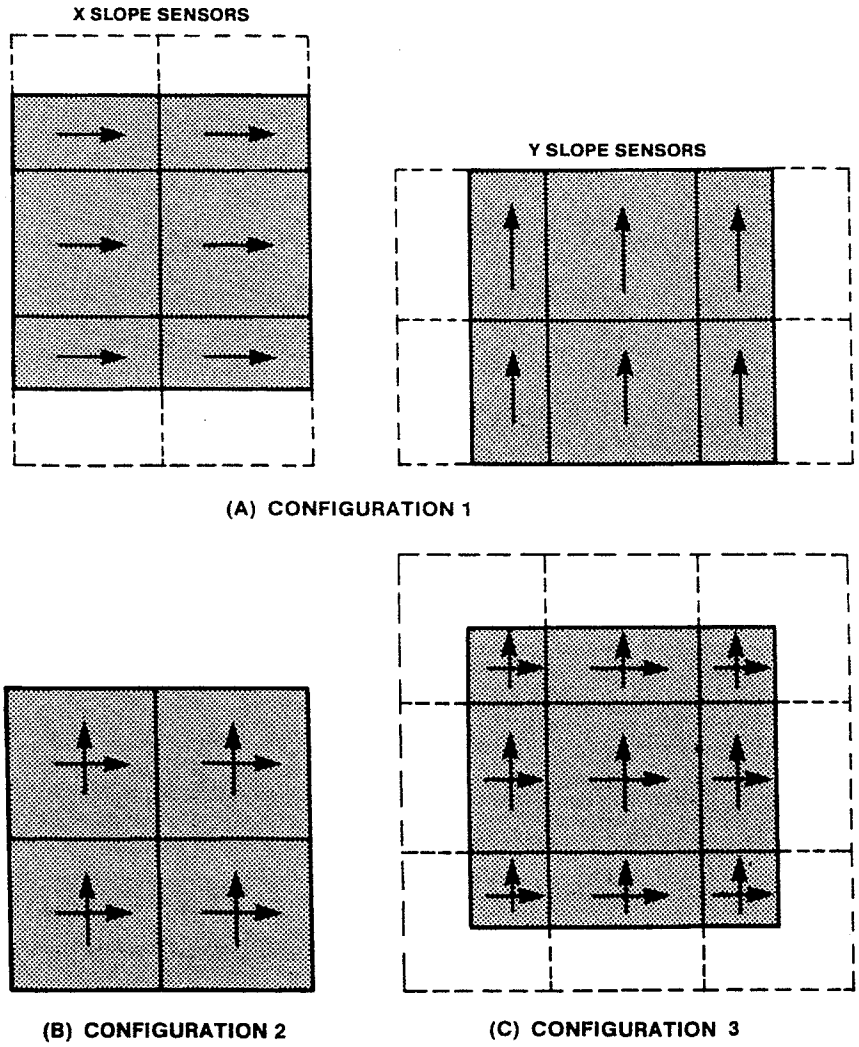



Figure 5. Sensor configurations for L = A/2.

In the second configuration, as shown in Figures 3 and 5b, x and y slopes are measured in the same subaperture. Here all sensors have the same signal strength and noise power.

The third configuration as shown in Figure 5c also measures x and y slopes in the same subapertures, but the array of sensors is shifted by half a subaperture in the x and y directions. Sensors at the edges of the aperture have their signal and noise powers halved, and for those in the corners of the aperture the signal and noise powers are quartered.

The wavefront being corrected is described by a Kolmogorov structure function

$$D(\bar{x}) = 6.884 \ |\bar{x}/r_0|^{5/3} \tag{33}$$

where $D(\bar{x})$ = structure function in radian squared of phase

$r_0$ = Fried's coherence length (m).

The mean square uncorrected phase error over the full aperture for this structure function is

$$\sigma_w^2 = 1.3103 \ (A/r_0)^{5/3} \tag{34}$$

where $\sigma_w^2$ = phase error ($rad^2$)

A = dimension of side of aperture (m)

The mean square slope measured over a sensor subaperture is

$$\sigma_s^2 = 6.4051 \ L^{-1/3} \ r_0^{-5/3} \ (rad^2 \ m^{-2}) \tag{35}$$

where L = side of sensor subaperture (m)

The photon noise on the measurement expressed in radians squared of phase difference is inversely proportional to the number of photons used in the measurement

$$\sigma_n^2 = K/BL^2 \ (rad^2) \tag{36}$$

where $\sigma_n^2$ = sensor noise ($rad^2$)

K = constant ($rad^2$ photons $m^{-2}$ $s^{-1}$)

B = wavefront irradiance (photons $m^{-2}$ $s^{-1}$)

The constant K will depend on the distribution of radiance over the source of the wavefront, the effective exposure time used in the measurement, the optical efficiency of the system, and the details of operation of the wavefront sensor.

Figure 6 shows the shape of the correction functions and the mean square residual error for an array of sensors in the second configuration with L = A/2. In this case, the four x and the four y correction functions have the same shape except for reflections and rotations. The mean square residual error represents the result of applying the corrections to a series of random wavefronts, squaring the residual phases after correction, and averaging over the series.

When the noise is very high, the correction function approaches a pure tilt, but the shape is unimportant because the magnitude of the correction becomes small. The residual error then approaches the uncorrected wavefront that is dominated by tilt, leading to an approximately quadratic surface for the mean square residual error, as seen in Figure 6b.

As the noise is reduced, the correction becomes more effective, reducing but not eliminating the tilt. The shape of the residual error, as shown in Figure 6d, is still predominantly quadratic, but the magnitude has been reduced. (Note that the scale is varied from case to case in Figure 6.)

In the limit of zero noise, the low spatial frequency terms are eliminated and the high-frequency terms dominate in the residual wavefront as in Figure 6h.

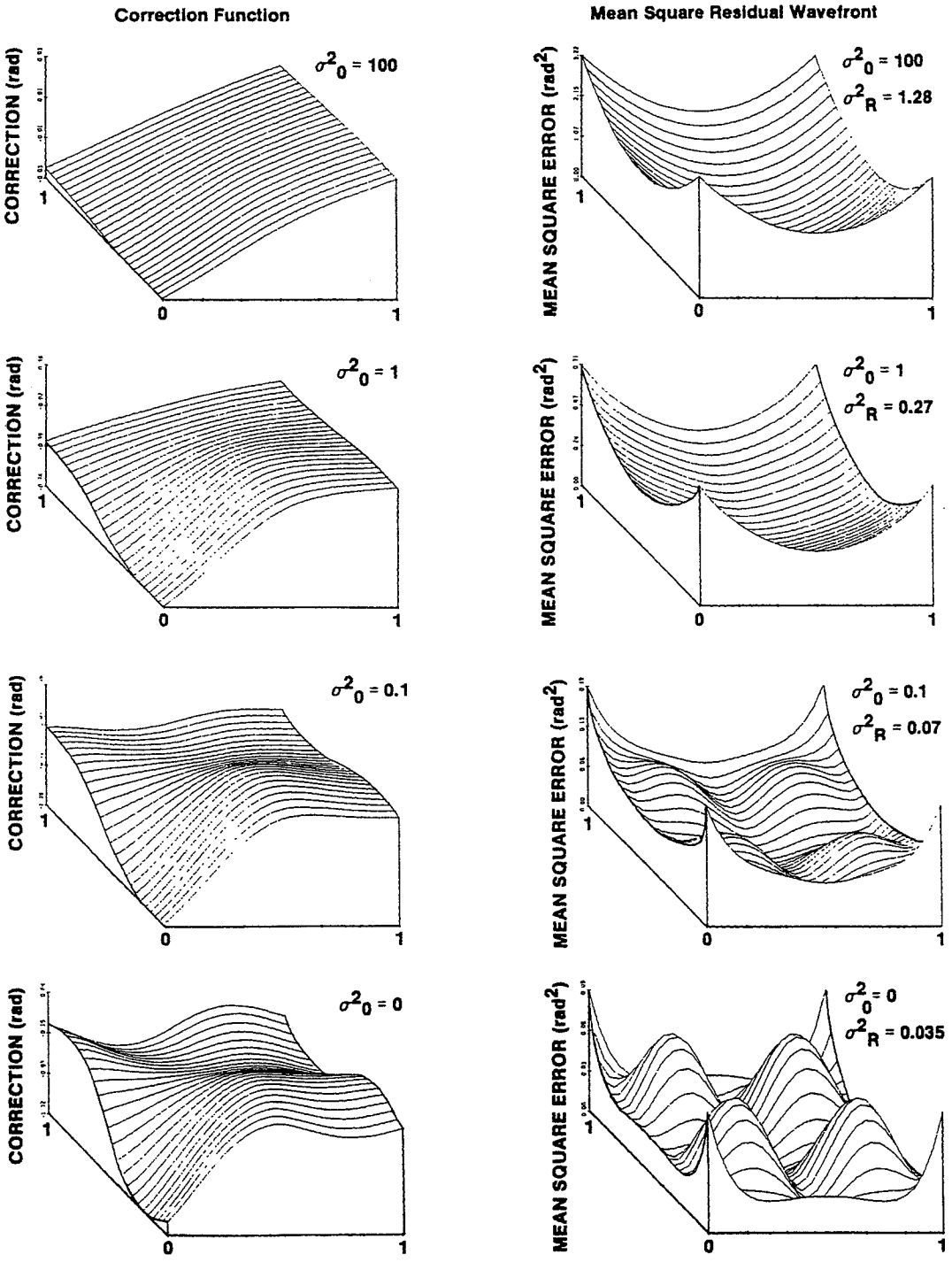



Figure 6. Performance versus noise level.

Figure 7 shows the effect on the mean square residual error of decreasing the subaperture size. The remaining error is predominantly at the spatial frequency of the sensor array, with magnitude proportional to $L^{5/3}$. This fitting error associated with the sensors is similar to the fitting error for an array of actuators that has been analyzed by Hudgin.[9]

The overall performance of the first configuration is as a function of noise level and subaperture size is shown by the solid curves in Figure 8. For photon noise levels above unity, the error is independent of the number of subapertures across the aperture, since at these levels only the tilt term is estimated with any reliability.

At low noise levels, the error depends only on the number of measurements. For a subaperture dimension equal to the side of the full aperture, the first configuration makes two measurements of x and two of y compared to one each for the second configuration. As the subaperture dimension is reduced, these edge effects become insignificant and the performance of the two configurations becomes identical.

The apparent difference in the performance of these configurations indicated by the analyses of Hudgin[3] and Fried[4] has been resolved by including the statistics of the wavefront phase as well as the photon noise in the treatment.

Figure 9 shows the performance for the first and third configurations. When the subaperture dimension is equal to the side of the aperture, the third configuration consists only of the four corners of the array, which is the same as a two-by-two array in the second configuration. As in the previous case, the edge and corner effects decrease as the subaperture size is reduced. Again, the differences in performance indicated in Southwell's analysis[8] disappear with a more realistic model and use of a priori knowledge of the wavefront statistics.

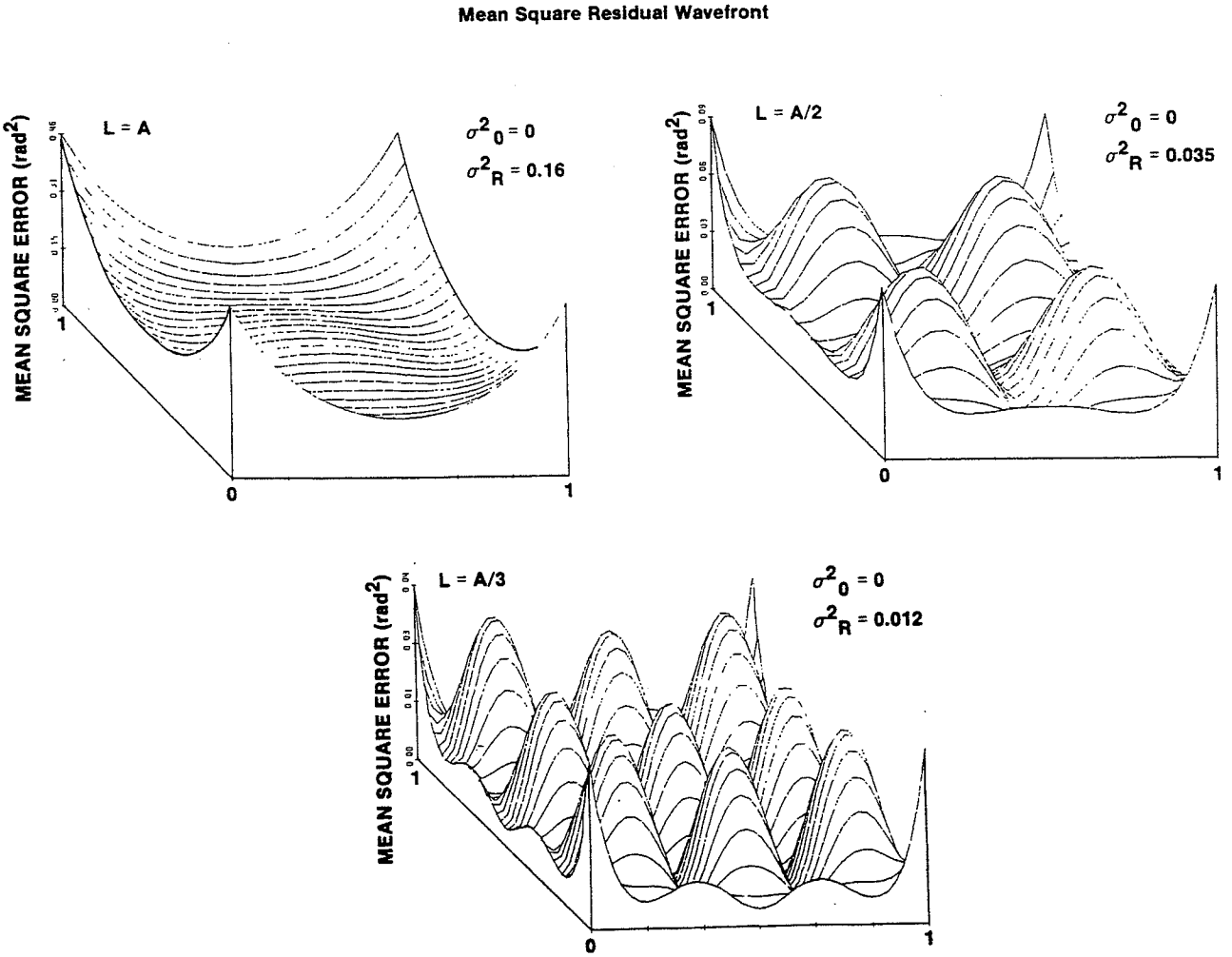



Figure 7. Performance versus subaperture size.

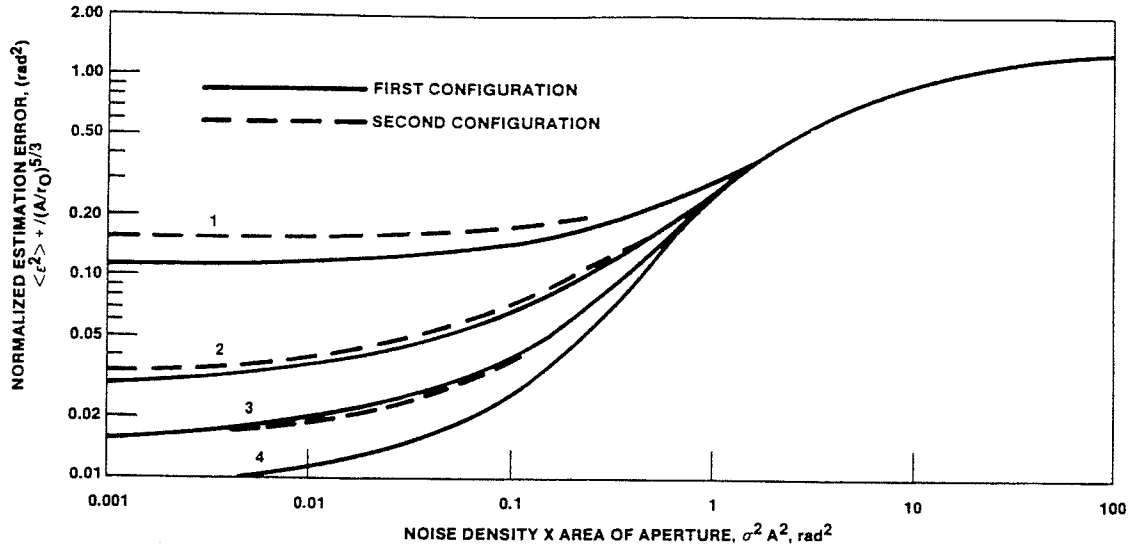



Figure 8. performance for first and second configurations.

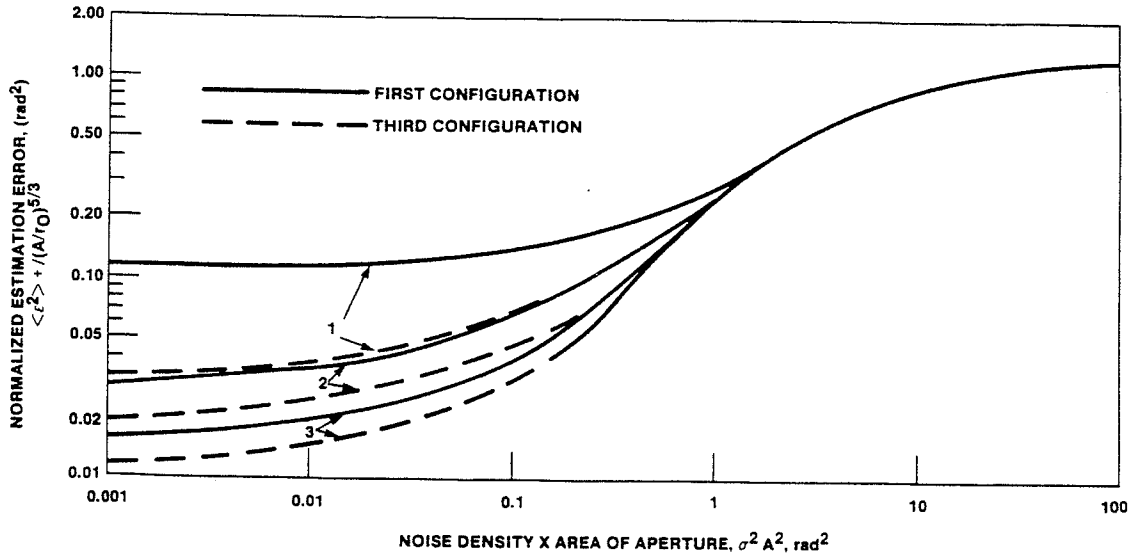



Figure 9. Performance for first and third configurations.

## References


1. Rimmer, M.P., "Methods for Evaluating Lateral Shear Interferograms," Appl. Opt., 13, 623-629 (1974).
2. Hardy, J.W., J.E. Lefebvre and C.L. Koliopoulos, "Real-Time Atmospheric Compensation," J. Opt. Soc. Am., 67, No. 3, 360-369 (1977).
3. Hudgin, R.H., "Wavefront Reconstruction for Compensated Imaging, J. Opt. Soc. Am., 67, No. 3, 375-378 (1977).
4. Fried, D.L., "Least Square Fitting a Wavefront Distortion Estimate to an Array of Phase-Difference Measurements," J. Opt. Soc. Am., 67, No. 3., 370-375 (1977).
5. Herrmann, J., "Least-squares wave front errors of minimum norm," J. Opt. Soc. Am., 70, No. 1, 28-35 (1980).
6. Wang, J.Y. and J.K. Markey, "Model Compensation of Atmospheric Turbulence Phase Distortion," J. Opt. Soc. Am., 68, No. 1, 78-88 (1978).
7. Wallner, E.P., "Optimal Wavefront Correction Using Slope Measurements," Itek Corporation, Memo of July 7, 1982.
8. Southwell, W.H., "Wavefront Estimation from Wavefront Slope Measurements, J. Opt. Soc. Am., 70, No. 8, 998-1005 (1980).
9. Hudgin, R.H. "Wavefront Compensation Error Due to Finite Corrector-Element Size, J. Opt. Soc. Am., 67, No. 3, 393-395 (1977).

# APPENDIX L

Reprinted with permission from *Journal of the Optical Society of America,* Vol. 73(12), pp. 1771-1776 (1983). 

# Optimal wave-front correction using slope measurements

Edward P. Wallner

*Optical Systems Division, Itek Corporation, 10 Maguire Road, Lexington, Massachusetts 02173*

Received July 26, 1982; revised manuscript received July 17, 1982

In adaptive optical systems that compensate for random wave-front disturbances, a wave front is measured and corrections are made to bring it to the desired shape. For most systems of this type, the local wave-front slope is first measured, the wave front is next reconstructed from the slope, and a correction is then fitted to the reconstructed wave front. Here a more realistic model of the wave-front measurements is used than in the previous literature, and wave-front estimation and correction are analyzed as a unified process rather than being treated as separate and independent processes. The optimum control law is derived for an arbitrary array of slope sensors and an arbitrary array of correctors. Application of this law is shown to produce improved results with noisy measurements. The residual error is shown to depend directly on the density of the slope measurements, but the sensitivity to the precise location of the measurements that was indicated in the earlier literature is not observed.

## INTRODUCTION

The wave-front correcting system treated here consists of a number of sensors, each of which measures the weighted average of wave-front slope over some region; a number of corrector actuators, each of which adds a corrective function to the wave-front phase; and a control law relating actuator displacements to sensor measurements. The correction will be imperfect because of the finite size and number of sensors, measurement noise (including photon noise in the sensors), the finite number of degrees of freedom in the corrector (whether distributed by zones or by modes), and possible deficiencies in the control law.

The approach taken to the control law in the literature has been first to reconstruct an estimate of the wave front from the slope data and then to fit the actuator functions to the estimate.[1] A number of papers have dealt with the problem of wave-front reconstruction from noisy data.[2-10] These papers generally treated the wave front as an array of discrete points and the measurements as phase differences between pairs of points, and only Ref. 10 included the effect of integrating wave-front slope over a finite area. The least-mean-squares estimate of the wave-front phase based on noisy measurements was derived, but no use was made in the process of any knowledge of the statistics of the wave front being measured. The latter information is useful even in the case of small measurement errors since there are always wave-front modes to which the sensor array is blind.

The problem of fitting a corrector to a wave front has also been analyzed by several authors, with particular application to wave fronts with the Kolmogorov spectrum typical of atmospheric turbulence.[11-15]

In practice, all these problems are interdependent, and the derivation of the optimal control law requires a unified treatment of the overall system. The present paper gives such a treatment for an arbitrary array of wave-front sensors, an arbitrary array of corrector functions, and an arbitrary distribution of uncorrected wave fronts. The control law that minimizes the mean-square residual error and the error itself are evaluated.

## SYSTEM DESCRIPTION AND ASSUMPTIONS

The overall correction system may be represented as a number of sensors that measure wave-front slope in the plane of an aperture, a corrector with a number of actuators acting in the same plane, and a control law connecting sensors to actuators. The aperture of the optical system can be described by a weighting function proportional to the intensity of light at the wave front. It is convenient to normalize this function so that

$$\int_{-\infty}^{\infty} d\mathbf{x}\, W_A(\mathbf{x}) = 1, \tag{1}$$

where $\mathbf{x}$ is the two-dimensional vector position in the aperture plane, $W_A(\mathbf{x})$ is the aperture weighting function, and $\int_{-\infty}^{\infty} d\mathbf{x}$ indicates integration over the entire aperture plane.

The uncorrected wave-front phase or optical path difference is described by the function $\psi(\mathbf{x})$. Because any uniform phase over the aperture does not affect optical performance, we can equivalently deal with the phase with aperture average removed, $\phi(\mathbf{x})$. For this function

$$\int_{-\infty}^{\infty} d\mathbf{x}\, W_A(\mathbf{x})\, \phi(\mathbf{x}) = 0, \tag{2}$$

where

$$\phi(\mathbf{x}) = \psi(\mathbf{x}) - \int_{-\infty}^{\infty} d\mathbf{x}'\, W_A(\mathbf{x}')\, \psi(\mathbf{x}')$$

and $\psi(\mathbf{x})$ is the input wave-front phase.

The output of each wave-front sensor is taken as the weighted average of the wave-front slope over some region of the aperture, which approximates the output of real wave-front slope sensors for small wave-front deviations and a fixed wave-front intensity pattern. The noise on the measurements is represented by a random additive function of position, which permits a convenient treatment of photon noise. The signal out of the $n$th sensor is then

$$s_n = \int_{-\infty}^{\infty} d\mathbf{x}\, W_{sn}(\mathbf{x})[\phi_n^s(\mathbf{x}) + v(\mathbf{x})], \tag{3}$$

where

$s_n$ is the signal from the $n$th sensor,
$W_{sn}(\mathbf{x})$ is the weighting function for the $n$th sensor,
$\phi_n^s(\mathbf{x})$ is the slope of wave front in the direction of sensitivity of $n$th sensor, and
$v(\mathbf{x})$ is the noise distribution over the aperture.

The signal can be expressed in terms of the wave-front phase by integrating Eq. (3) by parts:

$$s_n = \int_{-\infty}^{\infty} d\mathbf{x}[-W_{sn}^s(\mathbf{x})\phi(\mathbf{x}) + W_{sn}(\mathbf{x})v(\mathbf{x})], \tag{4}$$

where $W_{sn}^s(\mathbf{x})$ is the derivative of $W_{sn}(\mathbf{x})$ in the direction of slope sensitivity of the $n$th sensor. In Eq. (4) use has been made of the fact that $W_{sn}(\mathbf{x})$ goes to zero at infinity.

The noise is assumed to be unbiased and uncorrelated with the wave-front phase:

$$\langle v(\mathbf{x})\rangle = \langle v(\mathbf{x})\phi(\mathbf{x}')\rangle = 0. \tag{5}$$

The control law for the system generates a command to each actuator of the wave-front corrector based on all the sensor outputs. For a linear control law, the command to the $j$th actuator is

$$c_j = \sum_n M_{jn}s_n, \tag{6}$$

where $c_j$ is the command to the $j$th actuator and $M_{jn}$ is the weighting of $n$th sensor signal in $j$th actuator command.

Finally, the responses of the actuators are assumed to combine linearly to form the total wave-front correction,

$$\hat{\phi}(\mathbf{x}) = \sum_j c_j r_j(\mathbf{x}), \tag{7}$$

where $\hat{\phi}(\mathbf{x})$ is the total wave-front correction and $r_j(\mathbf{x})$ is the response of the $j$th actuator to a unit command.

## EVALUATION OF MEAN-SQUARE RESIDUAL ERROR

The residual wave-front error may be expressed by using Eq. (7) and expanded by using Eqs. (4) and (6):

$$\begin{aligned}\epsilon(\mathbf{x}) &= \hat{\phi}(\mathbf{x}) - \phi(\mathbf{x}) \\ &= \sum_j r_j(\mathbf{x}) \sum_n M_{jn}s_n - \phi(\mathbf{x}).\end{aligned} \tag{8}$$

The mean-square residual error at $\mathbf{x}$ is formed by using Eq. (8) and assuming that the sensor noise and the wave-front phase are uncorrelated:

$$\begin{aligned}\langle \epsilon^2(\mathbf{x})\rangle &= \sum_j \sum_{j'} \sum_n \sum_{n'} r_j(\mathbf{x})r_{j'}(\mathbf{x})M_{jn}M_{j'n'}\langle s_n s_{n'}\rangle \\ &\quad -2\sum_j \sum_n r_j(\mathbf{x})M_{jn}\langle s_n\phi(\mathbf{x})\rangle + \langle \phi^2(\mathbf{x})\rangle,\end{aligned} \tag{9}$$

where $\langle f\rangle$ is the expected value of $f$.

The overall measure of performance is the expected mean-square error averaged over the aperture:

$$\langle \epsilon_c^2\rangle = \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})\langle \epsilon^2(\mathbf{x})\rangle \tag{10}$$

$$\begin{aligned}&= \sum_j \sum_{j'} \sum_n \sum_{n'} M_{jn}M_{j'n'}\langle s_n s_{n'}\rangle \\ &\quad \times \int_{-\infty}^{\infty} d\mathbf{x}\, W_A(\mathbf{x})r_j(\mathbf{x})r_{j'}(\mathbf{x}) \\ &\quad -2\sum_j \sum_n M_{jn}\int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})r_j(\mathbf{x})\langle s_n\phi(\mathbf{x})\rangle \\ &\quad + \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})\langle \phi^2(\mathbf{x})\rangle.\end{aligned} \tag{11}$$

These expressions may be simplified by defining additional terms. The integral involving the product of two sensor functions is defined as the matrix $S_{nn'}$:

$$\begin{aligned}S_{nn'} &= \langle s_n s_{n'}\rangle \\ &= \int_{-\infty}^{\infty} d\mathbf{x}' \int_{-\infty}^{\infty} d\mathbf{x}''[W_{sn}^s(\mathbf{x}')W_{sn}^s(\mathbf{x}'')\langle \phi(\mathbf{x}')\phi(\mathbf{x}'')\rangle \\ &\quad + W_{sn}(\mathbf{x}')W_{sn'}(\mathbf{x}'')\langle v(\mathbf{x}')v(\mathbf{x}'')\rangle].\end{aligned} \tag{12}$$

The integral involving the product of two actuator response functions is defined as the matrix $R_{jj'}$:

$$R_{jj'} = \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})r_j(\mathbf{x})r_{j'}(\mathbf{x}). \tag{13}$$

The integral involving products of sensor and actuator response functions is defined as the matrix $A_{jn}$:

$$\begin{aligned}A_{jn} &= \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})r_j(\mathbf{x})\langle s_n\phi(\mathbf{x})\rangle \\ &= \int_{-\infty}^{\infty} d\mathbf{x} \int_{-\infty}^{\infty} d\mathbf{x}' W_A(\mathbf{x})r_j(\mathbf{x})W_{sn}^s(\mathbf{x}')\langle \phi(\mathbf{x})\phi(\mathbf{x}')\rangle.\end{aligned} \tag{14}$$

The average mean-square uncorrected error is defined as $\langle \epsilon_0^2\rangle$:

$$\langle \epsilon_0^2\rangle = \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x})\langle \phi^2(\mathbf{x})\rangle. \tag{15}$$

Substituting these definitions into Eq. (11) gives the simplified form

$$\langle \epsilon_c^2\rangle = \sum_j \sum_{j'} \sum_n \sum_{n'} M_{jn}M_{j'n'}S_{nn'}R_{jj'} - 2\sum_j \sum_n M_{jn}A_{jn} + \langle \epsilon_0^2\rangle. \tag{16}$$

## MINIMIZATION OF RESIDUAL ERROR

Equation (16) gives the average mean-square residual error for arbitrary weights in the control matrix $M_{jn}$. To optimize system performance we wish to choose the weights that minimize the error.

Differentiating Eq. (16) with respect to element $M_{j'n'}$ and equating to zero yields

$$\begin{aligned}\frac{d\langle \epsilon_c^2\rangle}{dM_{j'n'}} &= 2\sum_j \sum_n M_{jn}S_{nn'}R_{jj'} - 2A_{j'n'} \\ &= 0.\end{aligned} \tag{17}$$

By letting repeated indices indicate summation and making use of the symmetry of $R_{jj'}$, Eq. (17) may be written more concisely as

$$R_{j'j}M_{jn}S_{nn'} = A_{j'n'}. \tag{18}$$

The solution to this equation is the minimizing control matrix $M^*_{jn}$:

$$M^*_{jn} = R^{-1}_{jj'} A_{j'n'} S^{-1}_{n'n}. \tag{19}$$

The $R$ matrix will not be singular unless there are redundant actuators that allow the same correction to be made with different combinations of actuator commands. In that case the generalized inverse of the matrix may be used at the expense of additional computation, or the redundant functions can be eliminated with equivalent results.

Alternatively, a cost function depending on the actuator commands can be added to the quantity being minimized. If the cost rises more than linearly with the actuator command, the correction will be distributed over the redundant modes, and $R$ will not be singular. The $S$ matrix will not be singular, even with redundant sensors, if there is independent noise on each measurement, as will be the case in any real system.

The value of the minimum average mean-square residual error, $\langle \epsilon^2_r \rangle$, is found by substituting Eq. (19) in Eq. (16):

$$\begin{aligned} \langle \epsilon^2_r \rangle &= \langle \epsilon^2_0 \rangle - M^*_{jn} A_{jn} \\ &= \langle \epsilon^2_0 \rangle - R^{-1}_{jj'} A_{j'n'} S^{-1}_{nn'} A_{jn}. \end{aligned} \tag{20}$$

This expression, with Eqs. (12)–(14), gives the optimal performance for an arbitrary system and arbitrary wave-front statistics.

The overall performance will depend on the magnitude of the sensor noise relative to the expected signal that is due to fluctuations of the input wave front. If the noises in all sensors, $\sigma^2_d$, are equal and are large compared with the signal variance, the matrix $S^{-1}_{n'n}$ approaches $I/\sigma^2_d$, the correction term in Eq. (20) becomes small, and the residual error approaches the uncorrected error.

For small measurement errors, the residual error will approach a limit set by the inability of the finite arrays of sensors and actuators to fit the wave front perfectly.

In the transition region where the residual error is small compared with the uncorrected error but large compared with the fitting error, the performance will depend directly on the measurement errors and the error propagation of the control matrix.

## REPRESENTATION BY STRUCTURE FUNCTION

In the case of atmospheric turbulence, the statistics of the input wave front are defined in terms of a structure function,

$$D(\mathbf{x}, \mathbf{x}') = \langle [\psi(\mathbf{x}) - \psi(\mathbf{x}')]^2 \rangle. \tag{21}$$

The correlation of phases with averages removed required in the computation of $S$, $A$, and $\langle \epsilon^2_0 \rangle$ can be derived directly by using Eq. (21). The average removed phase is

$$\begin{aligned} \phi(\mathbf{x}) &= \psi(\mathbf{x}) - \int_{-\infty}^{\infty} d\mathbf{x}'' W_A(\mathbf{x}'') \psi(\mathbf{x}'') \\ &= \int_{-\infty}^{\infty} d\mathbf{x}'' W_A(\mathbf{x}'') [\psi(\mathbf{x}) - \psi(\mathbf{x}'')]. \end{aligned} \tag{22}$$

The correlation required is

$$\begin{aligned} \langle \phi(\mathbf{x}) \phi(\mathbf{x}') \rangle &= \int_{-\infty}^{\infty} d\mathbf{x}'' \int_{-\infty}^{\infty} d\mathbf{x}''' W_A(\mathbf{x}'') W_A(\mathbf{x}''') \\ &\quad \times \langle [\psi(\mathbf{x}) - \psi(\mathbf{x}'')][\psi(\mathbf{x}') - \psi(\mathbf{x}''')] \rangle \\ &= -\tfrac{1}{2} \int_{-\infty}^{\infty} d\mathbf{x}'' \int_{-\infty}^{\infty} d\mathbf{x}''' W_A(\mathbf{x}'') W_A(\mathbf{x}''') \\ &\quad \times [D(\mathbf{x}, \mathbf{x}') - D(\mathbf{x}, \mathbf{x}''') - D(\mathbf{x}'', \mathbf{x}') \\ &\quad + D(\mathbf{x}'', \mathbf{x}''')] \\ &= -\tfrac{1}{2} D(\mathbf{x}, \mathbf{x}') + g(\mathbf{x}) + g(\mathbf{x}') - a, \end{aligned} \tag{23}$$

where

$$g(\mathbf{x}) = \tfrac{1}{2} \int_{-\infty}^{\infty} d\mathbf{x}'' W_A(\mathbf{x}'') D(\mathbf{x}, \mathbf{x}''), \tag{24}$$

$$a = \tfrac{1}{2} \int_{-\infty}^{\infty} d\mathbf{x}'' \int_{-\infty}^{\infty} d\mathbf{x}''' W_A(\mathbf{x}'') W_A(\mathbf{x}''') D(\mathbf{x}'', \mathbf{x}'''). \tag{25}$$

Substituting into Eq. (15),

$$\begin{aligned} \langle \epsilon^2_0 \rangle &= \int_{-\infty}^{\infty} d\mathbf{x} W_A(\mathbf{x}) \langle \phi^2(\mathbf{x}) \rangle \\ &= a. \end{aligned} \tag{26}$$

In the evaluation of $S$ and $A$, all terms on the right-hand side of Eq. (23) except the first can be seen to integrate to zero, leaving

$$\begin{aligned} S_{nn'} = \int_{-\infty}^{\infty} d\mathbf{x}' \int_{-\infty}^{\infty} d\mathbf{x}'' [&-\tfrac{1}{2} W^s_{sn}(\mathbf{x}') W^s_{sn'}(\mathbf{x}'') D(\mathbf{x}', \mathbf{x}'') \\ &+ W_{sn}(\mathbf{x}') W_{sn'}(\mathbf{x}'') \langle v(\mathbf{x}') v(\mathbf{x}'') \rangle], \end{aligned} \tag{27}$$

$$A_{jn} = -\tfrac{1}{2} \int_{-\infty}^{\infty} d\mathbf{x} \int_{-\infty}^{\infty} d\mathbf{x}' W_A(\mathbf{x}) r_j(\mathbf{x}) W^s_{sn}(\mathbf{x}') D(\mathbf{x}, \mathbf{x}'). \tag{28}$$

Equations (26)–(28) may be used with Eqs. (19) and (20) to express the optimal control law and evaluate its performance in terms of the structure function of the wave front being corrected.

## PHOTON NOISE

In many systems of interest, the limit on performance is set by the noise that is due to photon statistics in the wave-front measurements. The photon noise for a single wave front can be represented by making $v(\mathbf{x})$ a spatially white-noise function. The correlation of the noise on two wave-front measurements depends in detail on the configuration of the wave-front sensor used. If the measurements are made on separate wave fronts, either in sequence or separated by a beam splitter, the noises will be uncorrelated, even though they refer to the same area in the aperture.

Measurements of orthogonal slopes made simultaneously on the same wave front will also have uncorrelated noises. To reflect these conditions the correlation of the noise functions may be written as

$$\langle v(\mathbf{x}) v(\mathbf{x}') \rangle = k_{nn'} \sigma^2_n \delta(\mathbf{x} - \mathbf{x}'), \tag{29}$$

where

$\sigma_n^2$ is the photon noise density,

$k_{nn'} = \mathbf{l}_n \cdot \mathbf{l}_{n'}$ for $n$ and $n'$ measured simultaneously on the same area of the same wave front but

$k_{nn'}$ is 0 otherwise,

$\mathbf{l}_n$ is a unit vector in the direction of slope sensitivity of the $n$th measurement, and

$\delta(\mathbf{x})$ is the Dirac delta function.

The noise term in Eq. 10 or 27 may then be written as

$$\int_{-\infty}^{\infty} d\mathbf{x}' \int_{-\infty}^{\infty} d\mathbf{x}'' W_{sn}(\mathbf{x}') W_{sn}(\mathbf{x}'') \langle v(\mathbf{x}') v(\mathbf{x}'') \rangle$$

$$= k_{nn'} \sigma_n^2 \int_{-\infty}^{\infty} d\mathbf{x}' W_{sn}(\mathbf{x}') W_{sn'}(\mathbf{x}'). \quad (30)$$

In most cases the basic measurements will be disjoint, and $k_{nn'}$ will be the Kronecker delta if each measurement is treated separately. It may be convenient to combine a number of the basic measurements before processing in order to reduce the order of the matrices involved in the computation of the control law. In this case the more general form of $k_{nn'}$ would be needed.

## CLOSED-LOOP RECONSTRUCTION

Most wave-front correction systems that have been implemented up to the present time have used a feedback-control loop in which the actuators are driven to null the wave-front sensor signals generated by the input wave front. Such a system is analyzed here for comparison with the optimum system derived above.

If a set of commands $c_j$ is sent to the actuators, the phase correction created will be given by Eq. (7). Letting repeated indices indicate summation, we may write this as

$$\hat{\phi}(\mathbf{x}) = c_j r_j(\mathbf{x}). \quad (31)$$

If this correction is subtracted from the input wave front, the signal from the $n$th sensor as given in Eq. (4) becomes

$$s_n' = s_n - \int_{-\infty}^{\infty} d\mathbf{x}[-W_{sn}^s(\mathbf{x})\hat{\phi}(\mathbf{x})]$$

$$= s_n - \int_{-\infty}^{\infty} d\mathbf{x}[-W_{sn}^s(\mathbf{x}) c_j r_j(\mathbf{x})]. \quad (32)$$

The integral on the right-hand side of Eq. (32) for a unit actuator command defines the matrix $P_{nj}$, which relates the $n$th sensor output to the $j$th actuator input:

$$P_{nj} = \int_{-\infty}^{\infty} d\mathbf{x}[-W_{sn}(\mathbf{x}) r_j(\mathbf{x})]. \quad (33)$$

Substituting in Eq. (32), we find that

$$s_n' = s_n - P_{nj} c_j. \quad (34)$$

In the closed-loop operation considered here, the command vector $c_j$ is chosen to drive the corrected signal vector $s_n'$ to zero. This requires that

$$c_j = P_{nj}^+ s_n, \quad (35)$$

where $P_{nj}^+$ is the generalized inverse of $P_{nj}$. The generalized inverse minimizes the sum of the squares of the residual signals and sets to zero any actuator modes not detected by the sensors.

The expected mean-square error averaged over the aperture is found by substituting the control matrix $P_{jn}^+$ for $M_{jn}$ in Eq. (16):

$$\langle \epsilon_c^2 \rangle = \langle \epsilon_0^2 \rangle - 2P_{nj}^+ A_{jn} + P_{nj}^+ P_{n'j'}^+ S_{nn'} R_{jj'}. \quad (36)$$

For a system with a given configuration of sensors and actuators, Eq. (36) allows the performance using simple closed-loop control to be compared with that obtained using optimal control.

## COMPUTATIONAL RESULTS

The theory derived here has been applied to a system consisting of a square aperture with uniform illumination and a square array of identical actuators with Gaussian response functions. The array is positioned with an actuator at each corner of the aperture, and the ratio of the side of the aperture to the actuator spacing, $N$, is varied in the computations. The actuator response falls to $1/e$ at the adjacent actuator.

The wave front being corrected is assumed to have the Kolmogorov structure function typifying atmospheric turbulence:

$$D(\bar{x}, \bar{x}') = 6.8839|(\bar{x} - \bar{x}')/r_0|^{5/3}, \quad (37)$$

where $D(\mathbf{x}, \mathbf{x}')$ is the phase-structure function, rad$^2$, and $r_0$ is Fried's coherence length, $m$. The mean-square uncorrected phase error over the full aperture for this structure function is

$$\langle \epsilon_0^2 \rangle = 1.3103(A/r_0)^{5/3}, \quad (38)$$

where $\langle \epsilon_0^2 \rangle$ is the uncorrected phase error, rad$^2$, and $A$ is a side of aperture $m$.

The wave-front sensor investigated averages the wave-front slope with uniform weight over a square subaperture of dimensions equal to the actuator spacing. Two different configurations of such sensors are considered.

In the first configuration, shown in Fig. 1, $X$ and $Y$ slopes are measured in displaced subapertures. This configuration is typically used in the shearing interferometer sensor and was treated by Rimmer[2] and Hudgin.[4] The mean-square slope measured over a square subaperture of side $L$ for the Kolmogorov structure function is

$$\sigma_s^2 = 6.4051 L^{-1/3} r_0^{-5/3}, \quad (39)$$

where $\sigma_s^2$ is the mean-square slope signal, rad$^2$ m$^{-2}$, and $L = A/N$ is a subaperture dimension $m$. The noise on this measurement is assumed to be pure photon noise for which the

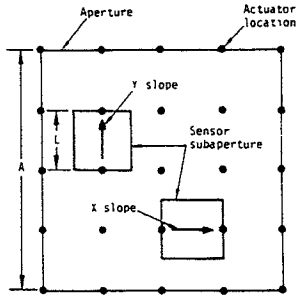



Fig. 1. Displaced subaperture wave-front sensor configuration.

mean-square slope error is inversely proportional to the number of photons used in the measurements.

In this configuration, a sensor subaperture located at an edge of the aperture parallel to the slope being measured will have only half of its area illuminated and will therefore have twice the mean-square slope error of a fully illuminated subaperture.

For a uniform illumination yielding $\rho$ photons per square meter, the noise will be

$$\sigma_n^2 = \frac{K}{M} = \frac{K}{\rho A_s},$$

where

$\sigma_n^2$ is the mean-square slope noise, rad$^2$ m$^{-2}$,
$K$ is a constant, rad$^2$ m$^{-2}$,
$M$ is the number of photoelectrons used,
$\rho$ is the photoelectron density $m^{-2}$, and
$A_s$ is the subaperture area $m^2$.

Figure 2 shows the residual error for this configuration as a function of photoelectron density and actuator spacing. The aperture size $A$ and $r_0$, where each is assumed to be 1 m, and a photon noise constant of 1204.6 were used. (This constant would apply to a shearing interferometer with optimum shear imaging a disk of 5-$\mu$rad diameter in light of 0.55-$\mu$m wavelength.)

The three regions of operation are easily distinguished in Fig. 2. For densities much less than 100, the measurement noise is so large that the *a priori* estimate of zero phase is more accurate than the measurement. For high densities, the error is determined by actuator and sensor spacing and is again independent of irradiance. This fitting error varies as $(A/Nr_0)^{5/3}$ when $N$ is sufficiently large that edge effects may be neglected.

In the transition region, where the photon error is small compared with the uncorrected error but large compared with the fitting error, the performance depends primarily on photoelectron density and depends relatively weakly on actuator spacing, in agreement with the results of Fried[3] and Hudgin.[4]

Figure 3 compares the performance of the optimal reconstructor with that of the closed-loop reconstructor for $N = 4$. At low photoelectron densities, the closed-loop reconstructor faithfully follows the noise input, leading to large errors. At high densities, the performance is close to that of the optimum, but the actuators are not fitted to the wave-front quite as well as possible. The optimal reconstructor thus improves performance at all illumination levels.

The second sensor configuration considered is shown in Fig. 4. Here the $x$ and $y$ slopes are measured in a common subaperture that has an actuator at each corner. This configuration characterizes the Hartmann sensor and was analyzed by Fried.[3]

The performance of the two configurations is compared in Fig. 5. The top dashed curve is for a single subaperture with four actuators, one at each corner. The fitting error in this case is slightly larger than for the displaced subaperture configuration with the same subaperture size, which uses twice as many slope measurements to link the four actuators. For larger numbers of actuators for which edge effects are reduced,

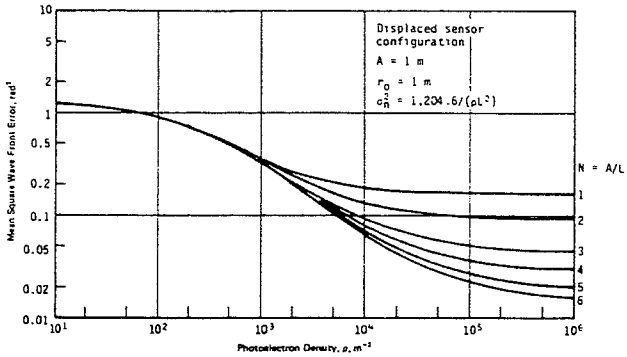



Fig. 2. Performance of displaced subaperture sensor configuration.

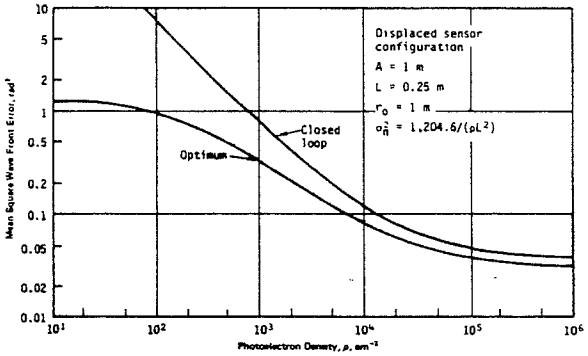



Fig. 3. Comparison of optimal and closed-loop reconstructors.

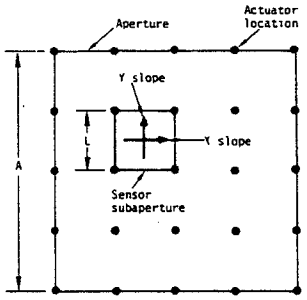



Fig. 4. Common subaperture wave-front sensor configuration.

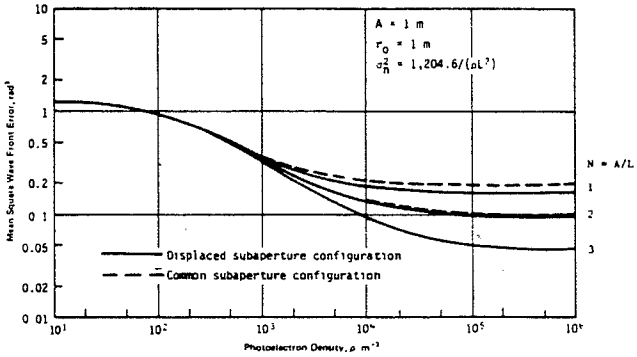



Fig. 5. Comparison of sensor configurations.

the performance of the two configurations is virtually indistinguishable. The difference in performance between these configurations as given by Fried[3] and Hudgin[4] does not appear in these results.

As was discussed by Herrmann,[8] when the wave-front reconstruction algorithm based on the average of the phases at the four nearest neighbors plus the average of the four connecting phase differences is used with the common subaperture configuration, two interleaved arrays of points are produced with an arbitrary phase difference between them. Fundamentally, the approach given here resolves the ambiguity by using the statistics of the wave fronts being estimated to imply the unobserved mode.

These results also contradict those of Southwell,[9] who shows a mean-square error for the common subaperture configuration approximately twice that of the displaced subaperture configuration. In his model, slopes were measured at discrete points with equal noise on all measurements, and the wave-front error was evaluated only over an array of discrete points.

With more-realistic sensors that average the slopes over finite subapertures, with a performance measure that evaluates the error over all the aperture, and with use of the statistics of the wave-front phase in the estimation process, the extreme sensitivity of performance to the precise sensor configuration reported by Southwell is not observed.

The same conclusion was reached in a separate paper[16] that dealt with the same sensor configurations but allowed actuator response functions to be chosen arbitrarily. Again, except for edge effects, performance was the same for the two configurations treated here as well as the configuration considered by Southwell in which $X$ and $Y$ slopes are measured in common subapertures centered at the actuator locations.

The results of the performance computations indicate that appreciable improvement in performance can be obtained by using the optimal reconstructor and that performance depends on sensor dimensions but not significantly on the position of the $X$ or $Y$ sensor array with respect to the actuator array.

## REFERENCES

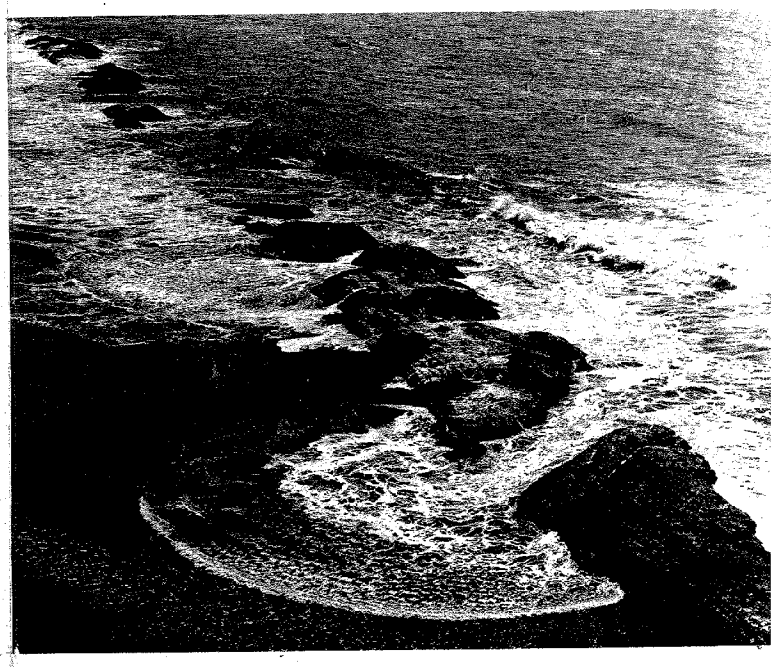

1. J. W. Hardy, J. E. Lefebvre, and C. L. Koliopoulos, "Real-time atmospheric compensation," J. Opt. Soc. Am. **67,** 360–369 (1977).
2. M. P. Rimmer, "Methods for evaluating lateral shear interferograms," Appl. Opt. **13,** 623–629 (1974).
3. D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," J. Opt. Soc. Am. **67,** 370–374 (1977).
4. R. H. Hudgin, "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am. **67,** 375–377 (1977).
5. B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," J. Opt. Soc. Am. **69,** 393–398 (1979).
6. R. J. Noll, "Phase estimates from slope-type wave-front sensors," J. Opt. Soc. Am. **68,** 139 (1978).
7. R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," J. Opt. Soc. Am. **69,** 972–977 (1979).
8. J. Herrmann, "Least squares wave-front errors of minimum norm," J. Opt. Soc. Am. **70,** 28–35 (1980).
9. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am. **70,** 998–1005 (1980).
10. J. Herrmann, "Cross-coupling and aliasing in modal wave-front estimation," J. Opt. Soc. Am. **71,** 989–992 (1981).
11. D. L. Fried, "Statistics for a geometric representation of wave-front distortion," J. Opt. Soc. Am. **55,** 1427–1435 (1965).
12. B. L. McGlamery and P. E. Silva (Visibility Laboratory, Scripps Institution of Oceanography, University of California, San Diego, California 92152), "A preliminary comparison of wave-front error measurement devices for use in compensated imaging systems" (unpublished report, March 1975).
13. D. L. Fried, "Required number of degrees-of-freedom for an adaptive optics system," Optical Sciences Consultants Rep. TR-191 (October 1975).
14. R. H. Hudgin, "Wave-front compensation error due to finite corrector-element size," J. Opt. Soc. Am. **67,** 393–395 (1977).
15. J. Y. Wang and J. K. Markey, "Modal compensation of atmospheric turbulence phase distortion," J. Opt. Soc. Am. **68,** 78–88 (1978).
16. E. P. Wallner, "Comparison of wave-front sensor configurations using optimal reconstruction and correction," Proc. Soc. Photo-Opt. Instrum. Eng. **351,** 42–53 (1982).

# APPENDIX M

MARSDEN • TROMBA

# VECTOR CALCULUS

SECOND EDITION



ISBN 0-7167-1244-X

FREEMAN

MARSDEN
TROMBA

VECTOR CALCULUS

SECOND EDITION

*Project Editor*: Larry Olsen
*Copy Editor*: Rebecca Stein
*Designer*: Sharon H. Smith
*Production Coordinator*: Bill Murdock
*Illustration Coordinator*: Audre Loverde
*Artist*: Evan Gillespie
*Compositor*: Syntax International
*Printer and Binder*: The Maple-Vail Book Manufacturing Group

The motion of waves is described by the equations of fluid mechanics, which are discussed on page 453. The notions of vector calculus can give insight into this and other complex phenomena. Photograph of Duxbury reef, Bolinas, California, by Nancy Williams.

**Library of Congress Cataloging in Publication Data**

Marsden, Jerrold E
Vector calculus.

Includes index.
1. Calculus. 2. Vector analysis. I. Tromba, Anthony, joint author. II. Title.
QA303.M338 1981 515'.63 80-24663
ISBN 0-7167-1244-X



AMS 1970 subject classifications: 26A60, 35-01

Printed in the United States of America

9 8 7 6 5 4 3 2

*To Ralph Abraham,*
*teacher and friend*

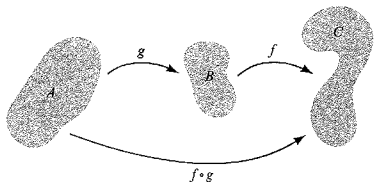



FIGURE 2.2.17
*The composition of $f$ on $g$.*

the composition of $g$ with $f$, or of $f$ on $g$, denoted by $f \circ g$, maps $A$ to $C$ by sending $\mathbf{x} \mapsto f(g(\mathbf{x}))$ (see Figure 2.2.17). For example, $\sin x^2$ is the composition of $x \mapsto x^2$ with $y \mapsto \sin y$. The reader should be familiar with this idea from one-variable calculus.

THEOREM 5. *Let $g: A \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $f: B \subset \mathbb{R}^m \to \mathbb{R}^p$. Suppose $g(A) \subset B$ so that $f \circ g$ is defined on $A$. If $g$ is continuous at $\mathbf{x}_0 \in A$ and $f$ is continuous at $\mathbf{y}_0 = g(\mathbf{x}_0)$, then $f \circ g$ is continuous at $\mathbf{x}_0$.*

The intuition behind this is easy; the formal proof in Appendix A follows a similar pattern. Intuitively we must show that as $\mathbf{x}$ gets close to $\mathbf{x}_0$, $f(g(\mathbf{x}))$ gets close to $f(g(\mathbf{x}_0))$. But as $\mathbf{x}$ gets close to $\mathbf{x}_0$, $g(\mathbf{x})$ gets close to $g(\mathbf{x}_0)$ (by continuity of $g$ at $\mathbf{x}_0$) and as $g(\mathbf{x})$ gets close to $g(\mathbf{x}_0)$, $f(g(\mathbf{x}))$ gets close to $f(g(\mathbf{x}_0))$ (by continuity of $f$ at $g(\mathbf{x}_0)$).

EXAMPLE 11. Let $f(x, y, z) = (x^2 + y^2 + z^2)^{30} + \sin z^3$. Show that $f$ is continuous.

Here we can write $f$ as a sum of two functions $(x^2 + y^2 + z^2)^{30}$ and $\sin z^3$, so it suffices to show that each is continuous. The first is the composite of $(x, y, z) \mapsto (x^2 + y^2 + z^2)$ with $u \mapsto u^{30}$ and the second is the composite of $(x, y, z) \mapsto z^3$ with $u \mapsto \sin u$, so we have continuity by Theorem 5.

**OPTIONAL**

## LIMITS IN TERMS OF ε's AND δ's

We shall now state a theorem giving a formulation of the notion of limit in terms of epsilons and deltas. This new formulation is quite useful; it is often taken as the *definition* of limit. It is another way of making precise the



**OPTIONAL (*Continued*)**

intuitive statement that "$f(\mathbf{x})$ is close to $\mathbf{b}$ when $\mathbf{x}$ is close to $\mathbf{x}_0$." To help understand this formulation, the reader should consider it with regards to each of the examples already presented.

THEOREM 6. *Let $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $\mathbf{x}_0$ be in $A$ or be a boundary point of $A$. Then $\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = \mathbf{b}$ if and only if for every number $\varepsilon > 0$ there is a $\delta > 0$ such that for any $\mathbf{x} \in A$ satisfying $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta$, we have $\|f(\mathbf{x}) - \mathbf{b}\| < \varepsilon$ (see Figure 2.2.18).*

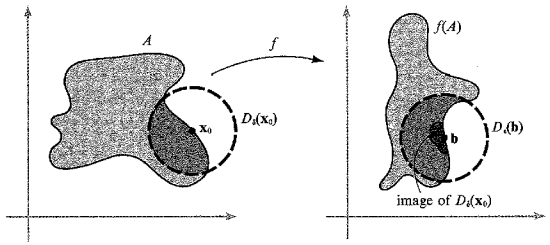



FIGURE 2.2.18
*The geometry of the ε-δ definition of limit.*

This theorem is proved in Appendix A.

To illustrate the mechanics of the epsilon-delta technique in Theorem 6, we consider the following examples.

EXAMPLE 12. Show that $\lim_{(x, y) \to (0, 0)} x = 0$.

Note that if $\delta > 0$, $\|(x, y) - (0, 0)\| = \sqrt{x^2 + y^2} < \delta$ implies $|x - 0| = |x| = \sqrt{x^2} \leq \sqrt{x^2 + y^2} < \delta$. Thus if $\|(x, y) - (0, 0)\| < \delta$, then $|x - 0|$ is also less than $\delta$. Given $\varepsilon > 0$ we must find a $\delta > 0$ (generally depending on $\varepsilon$) with the property that $0 < \|(x, y) - (0, 0)\| < \delta$ implies $|x - 0| < \varepsilon$. What are we to pick as our $\delta$? From the above calculation, we see that if we choose $\delta = \varepsilon$, then $\|(x, y) - (0, 0)\| < \delta$ implies $|x - 0| < \varepsilon$. This shows that $\lim_{(x, y) \to (0, 0)} x = 0$. We could have also chosen $\delta = \varepsilon/2$ or $\varepsilon/3$, but it suffices to find just one $\delta$ satisfying the requirements of the definition of limit.

**OPTIONAL (*Continued*)**

More generally, we can see in the same way that

$$\lim_{(x,\,y)\to(x_0,\,y_0)} x = x_0$$

This is just an epsilon-delta reformation of Example 4.

EXAMPLE 13. Consider the function

$$f(x, y) = \frac{\sin(x^2 + y^2)}{x^2 + y^2}$$

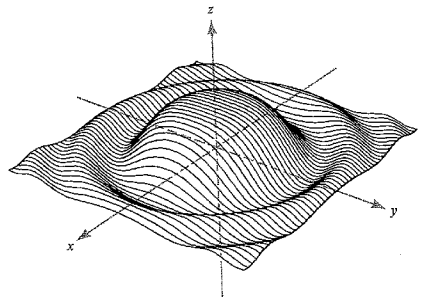
Even though $f$ is not defined at $(0, 0)$ we can ask whether $f(x, y)$ approaches some number as $(x, y)$ approaches $(0, 0)$. From elementary calculus we know that

$$\lim_{\alpha\to 0} \frac{\sin\alpha}{\alpha} = 1$$

Thus it is reasonable to guess that

$$\lim_{\mathbf{v}\to(0,\,0)} f(\mathbf{v}) = \lim_{\mathbf{v}\to(0,\,0)} \frac{\sin\|\mathbf{v}\|^2}{\|\mathbf{v}\|^2} = 1$$

Indeed, since $\lim_{\alpha\to 0} (\sin\alpha)/\alpha = 1$, given $\varepsilon > 0$ we can find a $\delta > 0$, with




$z = \dfrac{\sin(x^2 + y^2)}{(x^2 + y^2)}$ $\quad -3 \le x \le 3, -3 \le y \le 3$

FIGURE 2.2.19
*Computer-generated graph.*



**OPTIONAL (*Continued*)**

$1 > \delta > 0$, such that $0 < |\alpha| < \delta$ implies that $|(\sin\alpha)/\alpha - 1| < \varepsilon$. If $0 < \|\mathbf{v}\| < \delta$, then $0 < \|\mathbf{v}\|^2 < \delta^2 < \delta$, and therefore

$$|f(\mathbf{v}) - 1| = \left|\frac{\sin\|\mathbf{v}\|^2}{\|\mathbf{v}\|^2} - 1\right| < \varepsilon$$

Thus $\lim_{\mathbf{v}\to(0,\,0)} f(\mathbf{v}) = 1$. Indeed, if we plot $(\sin(x^2 + y^2))/(x^2 + y^2)$ on a computer, we get a graph that is well-behaved near $(0, 0)$ (Figure 2.2.19).

EXAMPLE 14. Show that

$$\lim_{(x,\,y)\to(0,\,0)} \frac{x^2}{\sqrt{x^2 + y^2}} = 0$$

We must show that $x^2/\sqrt{x^2 + y^2}$ is small when $(x, y)$ is close to the origin. To do this we need the following inequality:

$$0 \le \frac{x^2}{\sqrt{x^2 + y^2}} \le \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} \qquad \text{(since } y^2 \ge 0\text{)}$$
$$= \sqrt{x^2 + y^2}$$

Given $\varepsilon > 0$, choose $\delta = \varepsilon$. Then $\|(x, y) - (0, 0)\| = \|(x, y)\| = \sqrt{x^2 + y^2}$, so $\|(x, y) - (0, 0)\| < \delta$ implies that

$$\left|\frac{x^2}{\sqrt{x^2 + y^2}} - 0\right| = \frac{x^2}{\sqrt{x^2 + y^2}} \le \sqrt{x^2 + y^2} = \|(x, y) - (0, 0)\| < \delta = \varepsilon$$

Thus the conditions of Theorem 6 have been fulfilled and the limit verified.

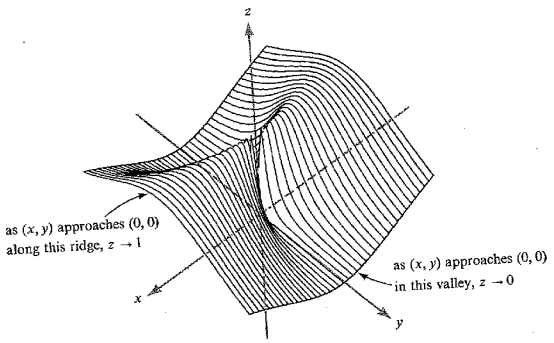
EXAMPLE 15. Does $\lim_{(x,\,y)\to(0,\,0)} x^2/(x^2 + y^2)$ exist?

If the limit exists, $x^2/(x^2 + y^2)$ should approach a definite value, say $a$, as $(x, y)$ gets near $(0, 0)$. In particular, if $(x, y)$ approaches zero along any given path then $x^2/(x^2 + y^2)$ should approach the limiting value $a$. If $(x, y)$ approaches $(0, 0)$ along the line $y = 0$, the limiting value is clearly 1 (just set $y = 0$ in the above expression to get $x^2/x^2 = 1$). If $(x, y)$ approaches $(0, 0)$ along the line $x = 0$, the limiting value is

$$\lim_{(x,\,y)\to(0,\,0)} \frac{0^2}{0^2 + y^2} = 0 \neq 1$$

Hence, $\lim_{(x,\,y)\to(0,\,0)} x^2/(x^2 + y^2)$ does not exist (see Figure 2.2.20).

**OPTIONAL (*Continued*)**




$$z = \frac{x^2}{x^2 + y^2} \quad -1 \le x \le 1, -1 \le y \le 1$$

FIGURE 2.2.20
*This function has no limit at* (0, 0).

EXAMPLE 16. Prove (see Figure 2.2.21)

$$\lim_{(x,\,y)\to(0,\,0)} \frac{x^2 y}{x^2 + y^2} = 0$$

Indeed, note that

$$\left|\frac{x^2 y}{x^2 + y^2}\right| \le \left|\frac{x^2 y}{x^2}\right| = |y|$$

Thus, given $\varepsilon > 0$, choose $\delta = \varepsilon$; so $0 < \|(x, y) - (0, 0)\| = \sqrt{x^2 + y^2} < \delta$ implies $|y| < \delta$, and thus

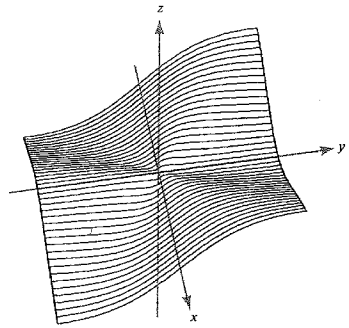
$$\left|\frac{x^2 y}{x^2 + y^2} - 0\right| < \varepsilon$$

For more examples of limits, see Appendix A.

Using the epsilon-delta notation, we are led to the following reformulation of the definition of continuity.



**OPTIONAL (*Continued*)**




$$z = \frac{x^2 y}{x^2 + y^2}$$

FIGURE 2.2.21
*This function has limit* 0 *at* (0, 0).

THEOREM 7. *Let $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ be given. Then $f$ is continuous at $\mathbf{x}_0 \in A$ if and only if for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that*

$$\mathbf{x} \in A \text{ and } \|\mathbf{x} - \mathbf{x}_0\| < \delta \quad \text{implies} \quad \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon$$

The proof is almost immediate. However, notice that in Theorem 6 we insisted that $0 < \|\mathbf{x} - \mathbf{x}_0\|$, that is, $\mathbf{x} \neq \mathbf{x}_0$. That is *not* imposed here; indeed, the conclusion of Theorem 7 is certainly valid when $\mathbf{x} = \mathbf{x}_0$, so there is no need to exclude this case. Here we do care about the value of $f$ at $\mathbf{x}_0$; we want $f$ at nearby points to be close to this value.

## EXERCISES

In the following exercises the reader may assume that the exponential function, sine, and cosine are continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

1. Show that the following subsets of the plane are open
   (a) $A = \{(x, y) | -1 < x < 1, -1 < y < 1\}$
   (b) $B = \{(x, y) | y > 0\}$

(c) $C = \{(x, y) | 2 < x^2 + y^2 < 4\}$

(d) $A \cup B \cup C$

(e) $D = \{(x, y) | x \neq 0 \quad \text{and} \quad y \neq 0\}$

2. Prove that for $\mathbf{x} \in \mathbb{R}^n$ and $s < t$, $D_s(\mathbf{x}) \subset D_t(\mathbf{x})$.

3. Prove that if $U$ and $V$ are neighborhoods of $\mathbf{x} \in \mathbb{R}^n$, then so are $U \cap V$ and $U \cup V$.

4. Prove that the boundary points of $]a, b[ \subset \mathbb{R}$ are the points $a$ and $b$.

*5. Use the $\varepsilon$-$\delta$ formulation of limits to prove that $x^2 \to 4$ as $x \to 2$. Give a shorter proof, using Theorem 3.

6. Compute the following limits

(a) $\lim_{(x, y) \to (0, 1)} x^3 y$

(b) $\lim_{(x, y) \to (0, 1)} e^x y$

(c) $\lim_{x \to 0} \dfrac{\sin^2 x}{x}$ (HINT: Recall that $\lim_{x \to 0} \dfrac{\sin x}{x} = 1$.)

(d) $\lim_{x \to 0} \dfrac{\sin^2 x}{x^2}$

7. Compute $\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x})$, if it exists, for the following cases:

(a) $f: \mathbb{R} \to \mathbb{R}, x \mapsto |x|, x_0 = 1$

(b) $f: \mathbb{R}^n \to \mathbb{R}, \mathbf{x} \mapsto \|\mathbf{x}\|$, arbitrary $\mathbf{x}_0$

(c) $f: \mathbb{R} \to \mathbb{R}^2, x \mapsto (x^2, e^x), x_0 = 1$

(d) $f: \mathbb{R}^2 \backslash \{(0, 0)\} \to \mathbb{R}^2, \ (x, y) \mapsto (\sin(x - y), \ e^{x(y+1)} - x - 1)/\|(x, y)\|, \ \mathbf{x}_0 = (0, 0)$

8. Let $A \subset \mathbb{R}^2$ be the unit disc $D_1(0, 0)$ with the point $\mathbf{x}_0 = (1, 0)$ added, and $f: A \to \mathbb{R}, \mathbf{x} \mapsto f(\mathbf{x})$ be the constant function $f(\mathbf{x}) = 1$. Show that $\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = 1$.

9. Let $f: \mathbb{R}^3 \to \mathbb{R}, f(x, y, z) = (x^2 + 3y^2)/(x + 1)$. Compute

$$\lim_{(x, y, z) \to (0, 0, 0)} f(x, y, z).$$

*10. Let $f: A \subset \mathbb{R}^n \to \mathbb{R}$ be given and let $\mathbf{x}_0$ be a boundary point of $A$. We say that $\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = \infty$ if for every $N > 0$ there is a $\delta > 0$ such that $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta$ implies $f(\mathbf{x}) > N$.

(a) Prove that $\lim_{x \to 1} (x - 1)^{-2} = \infty$.

Prove that $\lim_{x \to 0} 1/|x| = \infty$. Is it true that $\lim_{x \to 0} 1/x = \infty$?

(c) Prove that $\lim_{(x, y) \to (0, 0)} 1/(x^2 + y^2) = \infty$.

*11. Let $f: \mathbb{R} \to \mathbb{R}$ be a function. We write $\lim_{x \to b^-} f(x) = L$ and say that $L$ is the *left-hand limit* of $f$ at $b$, if for every $\varepsilon > 0$, there is a $\delta > 0$ such that $x < b$ and $0 < |x - b| < \delta$ implies $|f(x) - L| < \varepsilon$.

(a) Formulate a definition of *right-hand limit*, or $\lim_{x \to b^+} f(x)$.

(b) Find $\lim_{x \to 0^-} 1/(1 + e^{1/x})$ and $\lim_{x \to 0^+} 1/(1 + e^{1/x})$.

(c) Sketch the graph of $1/(1 + e^{1/x})$.



*12. Prove that there is a number $\delta > 0$ such that if $|a| < \delta$, then $|a^3 + 3a^2 + a| < 1/100$.

(b) Prove that there is a number $\delta > 0$ such that if $x^2 + y^2 < \delta^2$, then

$$|x^2 + y^2 + 3xy + 180xy^5| < 1/10{,}000.$$

13. Compute the following limits (review properties of the relevant functions if necessary).

(a) $\lim_{x \to 3} (x^2 - 3x + 5)$

(b) $\lim_{x \to 0} \sin x$

(c) $\lim_{h \to 0} \dfrac{(x + h)^2 - x^2}{h}$

(d) $\lim_{h \to 0} \dfrac{e^h - 1}{h}$ (HINT: Recall L'Hopital's rule.)

(e) $\lim_{x \to 0} \dfrac{\cos x - 1}{x^2}$

14. Compute the following limits.

(a) $\lim_{(x, y) \to (0, 0)} (x^2 + y^2 + 3)$

(b) $\lim_{(x, y) \to (0, 0)} \dfrac{xy}{x^2 + y^2 + 2}$

(c) $\lim_{(x, y) \to (0, 0)} \dfrac{e^{xy}}{x + 1}$

15. Prove that $\lim_{(x, y) \to (0, 0)} (\sin xy)/xy = 1$.

16. Show that the map $f: \mathbb{R} \to \mathbb{R}, x \mapsto x^2 e^x/(2 - \sin x)$ is continuous.

*17. Show that $f$ is continuous at $\mathbf{x}_0$ if and only if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \|f(\mathbf{x}) - f(\mathbf{x}_0)\| = 0$$

18. Show that $f: \mathbb{R} \to \mathbb{R}, x \mapsto (1 - x)^8 + \cos(1 + x^3)$ is continuous.

19. If $f: \mathbb{R}^n \to \mathbb{R}$ and $g: \mathbb{R}^n \to \mathbb{R}$ are continuous, show that the functions

$$f^2 g: \mathbb{R}^n \to \mathbb{R}, \mathbf{x} \mapsto (f(\mathbf{x}))^2 g(\mathbf{x})$$

and

$$f^2 + g: \mathbb{R}^n \to \mathbb{R}, \mathbf{x} \mapsto (f(\mathbf{x}))^2 + g(\mathbf{x})$$

are continuous.

20. Prove that $f: \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto ye^x + \sin x + (xy)^4$ is continuous.

21. Suppose $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$ and $\mathbf{x} \neq \mathbf{y}$. Show that there is a continuous function $f: \mathbb{R}^n \to \mathbb{R}$ with $f(\mathbf{x}) = 1$, $f(\mathbf{y}) = 0$, and $0 \leq f(\mathbf{z}) \leq 1$ for every $\mathbf{z}$ in $\mathbb{R}^n$.

*22. Show that $f: \mathbb{R}^n \to \mathbb{R}^m$ is continuous at all points if and only if the inverse image of every open set is open.

*23. Let $f: A \subset \mathbb{R}^n \to \mathbb{R}^m$ satisfy $\|f(\mathbf{x}) - f(\mathbf{y})\| \leq K\|\mathbf{x} - \mathbf{y}\|^\alpha$ for all $\mathbf{x}$ and $\mathbf{y}$ in $A$ for positive constants $K$ and $\alpha$. Show that $f$ is continuous. (Such functions are called *Hölder-continuous* or, if $\alpha = 1$, *Lipschitz-continuous*.)

## 7.1 GREEN'S THEOREM

Green's Theorem relates a line integral along a closed curve $C$ in the plane $\mathbb{R}^2$ to a double integral over the region enclosed by $C$. This important result will be generalized in the following sections to curves and surfaces in $\mathbb{R}^3$. We shall be referring to line integrals around curves that are the boundaries of elementary regions of type 1, 2 or 3 (see Section 5.3). To understand the ideas in this section you may need to refer to Section 6.2.

A simple closed curve $C$ that is the boundary of a region of type 1, 2, or 3 has two orientations—counterclockwise (positive) and clockwise (negative). We denote $C$ with the counterclockwise orientation as $C^+$, and with the clockwise orientation as $C^-$ (Figure 7.1.1).

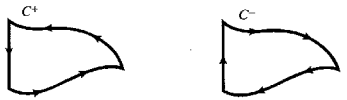



FIGURE 7.1.1
*Positive orientation of $C$ (a) and negative orientation of $C$ (b).*

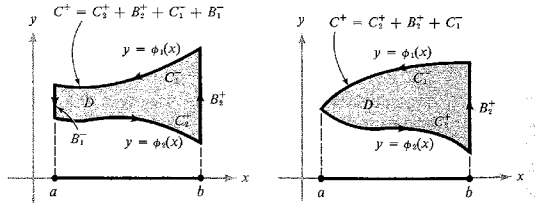



FIGURE 7.1.2
*Two examples showing how to break the positively oriented boundary of a region $D$ of type 1 into oriented components.*



The boundary $C$ of a region of type 1 can be decomposed into top and bottom portions, $C_1$ and $C_2$, and (if applicable) left and right vertical portions, $B_1$ and $B_2$. Then we write, following Figure 7.1.2,

$$C^+ = C_2^+ + B_2^+ + C_1^- + B_1^-$$

where the pluses denote the curves oriented in the direction of left to right or bottom to top, and the minuses denote the curves oriented from right to left or from top to bottom.

We may make a similar decomposition of the boundary of a region of type 2 into left and right portions, and upper and lower horizontal portions (if applicable) (Figure 7.1.3).

The boundary of a region of type 3 has two decompositions—one into upper and lower halves, the other into left and right halves.

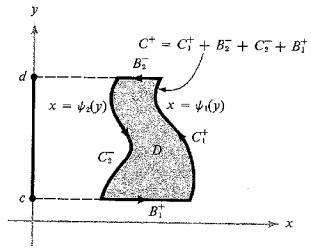



FIGURE 7.1.3
*An example showing how to break the positively oriented boundary of a region $D$ of type 2 into oriented components.*

We shall now prove two lemmas in preparation for Green's Theorem.

LEMMA 1. *Let $D$ be a region of type 1 and let $C$ be its boundary. Suppose $P: D \to \mathbb{R}$ is $C^1$. Then*

$$\int_{C^+} P\,dx = -\int_D \frac{\partial P}{\partial y}\,dx\,dy$$

*(The left-hand side denotes the line integral $\int_{C^+} P\,dx + Q\,dy + R\,dz$ where $Q = 0$ and $R = 0$.)*

*Proof.* Suppose $D$ is described by

$$a \le x \le b, \qquad \phi_2(x) \le y \le \phi_1(x)$$

We decompose $C^+$ by writing $C^+ = C_2^+ + B_2^+ + C_1^- + B_1^-$ (see Figure 7.1.2). By Fubini's Theorem, we may evaluate the double integral as an iterated integral (see p. 282)

$$\int_D \frac{\partial P}{\partial y}(x, y)\, dx\, dy = \int_a^b \int_{\phi_2(x)}^{\phi_1(x)} \frac{\partial P}{\partial y}(x, y)\, dy\, dx$$

$$= \int_a^b [P(x, \phi_1(x)) - P(x, \phi_2(x))]\, dx$$

However, since $C_2^+$ can be parametrized by $x \mapsto (x, \phi_2(x))$, $a \le x \le b$, and $C_1^+$ can be parametrized by $x \mapsto (x, \phi_1(x))$, $a \le x \le b$, we have

$$\int_a^b P(x, \phi_2(x))\, dx = \int_{C_2^+} P(x, y)\, dx$$

and

$$\int_a^b P(x, \phi_1(x))\, dx = \int_{C_1^+} P(x, y)\, dx$$

Thus by reversing orientations

$$-\int_a^b P(x, \phi_1(x))\, dx = \int_{C_1^-} P(x, y)\, dx$$

Hence

$$\int_D \frac{\partial P}{\partial y}\, dx\, dy = -\int_{C_2^+} P\, dx - \int_{C_1^-} P\, dx$$

Since $x$ is constant on $B_2^+$ and $B_1^-$ we have

$$\int_{B_2^+} P\, dx = 0 = \int_{B_1^-} P\, dx$$

and so

$$\int_{C^+} P\, dx = \int_{C_2^+} P\, dx + \int_{B_2^+} P\, dx + \int_{C_1^-} P\, dx + \int_{B_1^-} P\, dx$$

$$= \int_{C_2^+} P\, dx + \int_{C_1^-} P\, dx$$

Thus

$$\int_D \frac{\partial P}{\partial y}\, dx\, dy = -\int_{C_2^+} P\, dx - \int_{C^-} P\, dx = -\int_{C^+} P\, dx \quad \blacksquare$$

We now prove the analogous lemma with the roles of $x$ and $y$ interchanged.

LEMMA 2. *Let $D$ be a region of type 2 with boundary $C$. Then if $Q: D \to \mathbb{R}$ is $C^1$,*

$$\int_{C^+} Q\, dy = \int_D \frac{\partial Q}{\partial x}\, dx\, dy$$



The negative sign does not occur here because reversing the role of $x$ and $y$ corresponds to a change of orientation for the plane.

*Proof.* Suppose $D$ is given by

$$\psi_2(y) \le x \le \psi_1(y), \qquad c \le y \le d$$

Using the notation of Figure 7.1.3 we have

$$\int_{C^+} Q\, dy = \int_{C_2^- + B_1^+ + C_1^+ + B_2^-} Q\, dy = \int_{C_1^+} Q\, dy + \int_{C_2^-} Q\, dy$$

where $C_1^+$ is the curve parametrized by $y \mapsto (\psi_1(y), y)$, $c \le y \le d$, and $C_2^+$ is the curve $y \mapsto (\psi_2(y), y)$, $c \le y \le d$. Applying Fubini's Theorem, we obtain

$$\int_D \frac{\partial Q}{\partial x}\, dx\, dy = \int_c^d \int_{\psi_2(y)}^{\psi_1(y)} \frac{\partial Q}{\partial x}\, dx\, dy$$

$$= \int_c^d [Q(\psi_1(y), y) - Q(\psi_2(y), y)]\, dy$$

$$= \int_{C_1^+} Q\, dy - \int_{C_2^+} Q\, dy = \int_{C_1^+} Q\, dy + \int_{C_2^-} Q\, dy$$

$$= \int_{C^+} Q\, dy \quad \blacksquare$$

Adding the results of Lemmas 1 and 2 proves the following important theorem.

THEOREM 1 (GREEN'S THEOREM). *Let $D$ be a region of type 3 and let $C$ be its boundary. Suppose $P: D \to \mathbb{R}$ and $Q: D \to \mathbb{R}$ are $C^1$. Then*

$$\int_{C^+} P\, dx + Q\, dy = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy$$

The correct (positive) orientation for the boundary curves of region D can be remembered by the following device: "*If you walk along the curve C with the correct orientation, the region D will be on your left.*" See Figure 7.1.4.

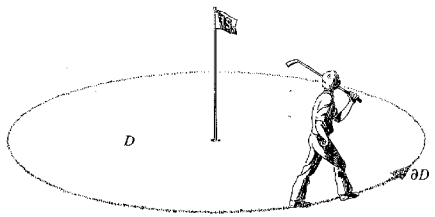



FIGURE 7.1.4
*The correct orientation for the boundary of a region D.*

Green's Theorem actually applies to any "decent" region in $\mathbb{R}^2$. In Exercise 8 we indicate a generalization of Green's theorem for regions that are not type 3, but that can be broken up into pieces, each of which is of type 3. An example is shown in Figure 7.1.5. The region $D$ is an annulus; its boundary consists of two curves $C = C_1 + C_2$ with the indicated orientations. (Note that for the inner region the correct orientation is *clockwise*; the device on p. 407 still works for remembering the orientation.) If Theorem 1 is applied to each of the regions $D_1$, $D_2$, $D_3$, and $D_4$ and the results summed, the equality of Green's Theorem will be obtained for $D$ and its boundary curve $C$.

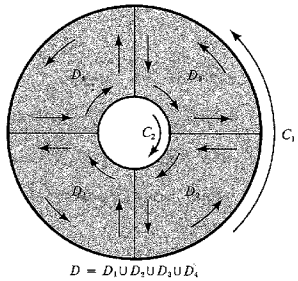



FIGURE 7.1.5
*Green's Theorem applies to* $D = D_1 \cup D_2 \cup D_3 \cup D_4$.

Let us use the notation $\partial D$ for the oriented curve $C^+$, that is, the boundary curve of $D$ oriented in the correct sense as described by the device on p. 407. Then we may write Green's Theorem as

$$\int_{\partial D} P\,dx + Q\,dy = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy$$

Green's Theorem is very useful because it relates a line integral around the boundary of a region to an area integral over the interior of the region, and in many cases it is easier to evaluate the line integral than the area integral. For example, if we know that $P$ vanishes on the boundary, we can immediately conclude that $\int_D (\partial P/\partial y)\,dx\,dy = 0$ even though $\partial P/\partial y$ need not vanish on the interior. (Can you construct such a $P$ on the unit square?)

EXAMPLE 1. Verify Green's Theorem for $P(x, y) = x$ and $Q(x, y) = xy$, where $D$ is the unit disc $x^2 + y^2 \le 1$.



We can evaluate both sides in Green's Theorem directly. The boundary of $D$ is the unit circle parametrized by $x = \cos t$, $y = \sin t$, $0 \le t \le 2\pi$, so

$$\int_{\partial D} P\,dx + Q\,dy = \int_0^{2\pi} [(\cos t)(-\sin t) + \cos t \sin t \cos t]\,dt$$
$$= \left[\frac{\cos^2 t}{2}\right]_0^{2\pi} + \left[-\frac{\cos^3 t}{3}\right]_0^{2\pi} = 0$$

On the other hand

$$\int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy = \int_D y\,dx\,dy$$

which is zero also. Thus Green's Theorem is verified in this case.

We can use Green's Theorem to obtain a formula for the area of a region bounded by a simple closed curve.

THEOREM 2. *If $C$ is a simple closed curve that bounds a region to which Green's Theorem applies, then the area of the region $D$ bounded by $C$ is*

$$A = \frac{1}{2}\int_{\partial D} x\,dy - y\,dx$$

*Proof.* Let $P(x, y) = -y$, $Q(x, y) = x$; then by Green's Theorem we have

$$\frac{1}{2}\int_{\partial D} x\,dy - y\,dx = \frac{1}{2}\int_D \left(\frac{\partial x}{\partial x} - \frac{\partial(-y)}{\partial y}\right) dx\,dy$$
$$= \int_D dx\,dy = A \quad \blacksquare$$

EXAMPLE 2. The area of the region enclosed by the hypocycloid $x^{2/3} + y^{2/3} = a^{2/3}$ can be computed using the parametrization

$$x = a\cos^3\theta, \qquad y = a\sin^3\theta, \qquad 0 \le \theta \le 2\pi$$

(see Figure 7.1.6). Thus

$$A = \frac{1}{2}\int_{\partial D} x\,dy - y\,dx$$

$$= \frac{1}{2}\int_0^{2\pi} [(a\cos^3\theta)(3a\sin^2\theta\cos\theta) - (a\sin^3\theta)(-3a\cos^2\theta\sin\theta)]\,d\theta$$

$$= \frac{3}{2}a^2\int_0^{2\pi} [\sin^2\theta\cos^4\theta + \cos^2\theta\sin^4\theta]\,d\theta = \frac{3}{2}a^2\int_0^{2\pi} \sin^2\theta\cos^2\theta\,d\theta$$

$$= \frac{3}{8}a^2\int_0^{2\pi} \sin^2 2\theta\,d\theta = \frac{3}{8}a^2\int_0^{2\pi}\left[\frac{1-\cos 4\theta}{2}\right] d\theta$$

$$= \frac{3}{16}a^2\int_0^{2\pi} d\theta - \frac{3}{16}a^2\int_0^{2\pi}\cos 4\theta\,d\theta = \frac{3}{8}\pi a^2$$

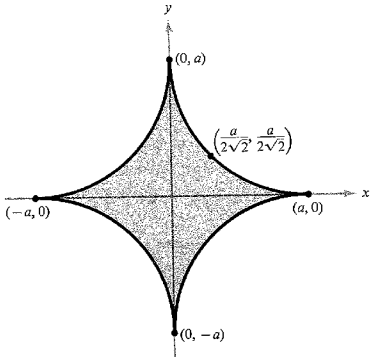



FIGURE 7.1.6
*The hypocycloid* $x = a\cos^3\theta$, $y = a\sin^3\theta$, $0 \le \theta \le 2\pi$.

The statement of Green's Theorem contained in Theorem 1 is not the form that we shall generalize. We can rewrite the theorem neatly in the language of vector fields.

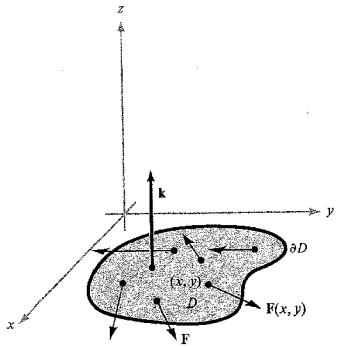



FIGURE 7.1.7
*The elements of the vector form of Green's Theorem.*



THEOREM 3 (VECTOR FORM OF GREEN'S THEOREM). *Let $D \subset \mathbb{R}^2$ be a region of type 3 and let $\partial D$ be its boundary (oriented counterclockwise). Let $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ be a $C^1$ vector field on $D$. Then*

$$\int_{\partial D} \mathbf{F} \cdot d\mathbf{s} = \int_D (\operatorname{curl} \mathbf{F}) \cdot \mathbf{k}\, dA = \int_D (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dA$$

(*see Figure 7.1.7*).

This result follows easily from Theorem 1 after we interpret the various symbols. We shall ask the reader to supply the details in Exercise 14.

EXAMPLE 3. Let $\mathbf{F} = (xy^2, y + x)$. Integrate $(\nabla \times \mathbf{F}) \cdot \mathbf{k}$ over the region in the first quadrant bounded by the curves $y = x^2$ and $y = x$.

*Method 1.* Here we compute

$$\nabla \times \mathbf{F} = \left(0, 0, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right) = (1 - 2xy)\mathbf{k}.$$

Thus $(\nabla \times \mathbf{F}) \cdot \mathbf{k} = 1 - 2xy$. This can be integrated over the given region $D$ (see Figure 7.1.8) using an iterated integral as follows:

$$\begin{aligned}\iint_D (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dx\, dy &= \int_0^1 \int_{x^2}^{x} (1 - 2xy)\, dy\, dx \\ &= \int_0^1 \left[y - xy^2\right]_{x^2}^{x} dx \\ &= \int_0^1 \left[x - x^3 - x^2 + x^5\right] dx = \tfrac{1}{2} - \tfrac{1}{4} - \tfrac{1}{3} + \tfrac{1}{6} = \tfrac{1}{12}\end{aligned}$$

*Method 2.* Here we use Theorem 3 to obtain

$$\iint_D (\nabla \times \mathbf{F}) \cdot \mathbf{k}\, dx\, dy = \int_{\partial D} \mathbf{F} \cdot d\mathbf{s}$$

The line integral of $\mathbf{F}$ along the curve $y = x$ from left to right is

$$\int_0^1 F_1\, dx + F_2\, dy = \int_0^1 (x^3 + 2x)\, dx = \tfrac{1}{4} + 1 = \tfrac{5}{4}$$

Along the curve $y = x^2$ we get

$$\int_0^1 F_1\, dx + F_2\, dy = \int_0^1 (x^5\, dx + (x + x^2)(2x\, dx) = \tfrac{1}{6} + \tfrac{2}{3} + \tfrac{1}{2} = \tfrac{4}{3}$$

Thus, remembering that the integral along $y = x$ is to be taken from right to left, as in Figure 7.1.8,

$$\int_{\partial D} \mathbf{F} \cdot d\mathbf{s} = \tfrac{4}{3} - \tfrac{5}{4} = \tfrac{1}{12}$$

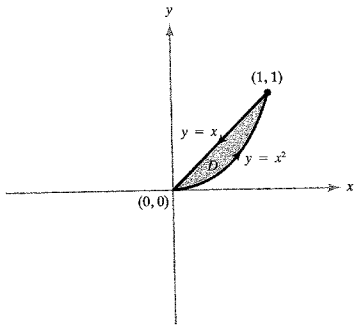



FIGURE 7.1.8
*The region bounded by the curves $y = x^2$ and $y = x$.*

There is yet another form of Green's Theorem that is capable of being generalized to $\mathbb{R}^3$.

THEOREM 4 (DIVERGENCE THEOREM IN THE PLANE). *Let $D \subset \mathbb{R}^2$ be a region of type 3 and let $\partial D$ be its boundary. Let $\mathbf{n}$ denote the outward unit normal to $\partial D$, which is given by*

$$\mathbf{n} = (y'(t), -x'(t))/\sqrt{(x'(t))^2 + (y'(t))^2}$$

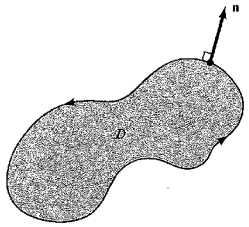



FIGURE 7.1.9
**n** *is the outward unit normal to $\partial D$.*



*if $\sigma: [a, b] \to \mathbb{R}^2$, $t \mapsto \sigma(t) = (x(t), y(t))$ is a positively oriented parametrization of $\partial D$ (see Figure 7.1.9).*

*Let $\mathbf{F} = P\mathbf{i} + Q\mathbf{j}$ be a $C^1$ vector field on $D$. Then*

$$\int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds = \int_D \operatorname{div} \mathbf{F}\, dA$$

*Proof.* Since $\sigma'(t) = (x'(t), y'(t))$ is tangent to $\partial D$ it is clear that $\mathbf{n} \cdot \sigma' = 0$, so $\mathbf{n}$ is normal to the boundary. The sign of $\mathbf{n}$ is chosen to make it correspond to the *outward* (rather than the inward) direction. By definition of the path integral (see Section 6.1)

$$\begin{aligned}\int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds &= \int_a^b \left(\frac{P(x(t), y(t))y'(t) - Q(x(t), y(t))x'(t)}{\sqrt{(x'(t))^2 + (y'(t))^2}}\right)\sqrt{(x'(t))^2 + (y'(t))^2}\, dt \\ &= \int_a^b [P(x(t), y(t))\, y'(t) - Q(x(t), y(t))x'(t)]\, dt \\ &= \int_{\partial D} P\, dy - Q\, dx\end{aligned}$$

By Green's Theorem, this equals

$$\int_D \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\, dy = \int_D \operatorname{div} \mathbf{F}\, dA \quad \blacksquare$$

EXAMPLE 4. Let $\mathbf{F} = y^3\mathbf{i} + x^5\mathbf{j}$. Compute the integral of the normal component of $\mathbf{F}$ around the unit square.

This can be done by the Divergence Theorem. Indeed,

$$\int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds = \int_D \operatorname{div} \mathbf{F}\, dA$$

But div $\mathbf{F} = 0$, so the integral is zero.

## EXERCISES

1. Evaluate $\int_C y\, dx - x\, dy$ where $C$ is the boundary of the square $[-1, 1] \times [-1, 1]$ oriented in the counterclockwise direction (use Green's Theorem).
2. Find the area of the disc $D$ of radius $R$ using Green's Theorem.
3. Verify Green's Theorem for the disc $D$ with center $(0, 0)$ and radius $R$ and the functions:
   (a) $P(x, y) = xy^2$, $Q(x, y) = -yx^2$;
   (b) $P(x, y) = x + y$, $Q(x, y) = y$; and
   (c) $P(x, y) = xy = Q(x, y)$.
4. Using the Divergence Theorem show that $\int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds = 0$ where $\mathbf{F}(x, y) = y\mathbf{i} - x\mathbf{j}$ and $D$ is the unit disc. Verify this directly.

## OPTIONAL (*Continued*)

Now

$$dx \wedge (dx \wedge dy) = (dx \wedge dx) \wedge dy = 0 \wedge dy = 0$$
$$dy \wedge (dx \wedge dy) = -dy \wedge (dy \wedge dx)$$
$$= -(dy \wedge dy) \wedge dx = 0 \wedge dx = 0$$

and

$$dz \wedge (dx \wedge dy) = (-1)^2(dx \wedge dy) \wedge dz = dx\,dy\,dz$$

Consequently

$$d(F\,dx\,dy) = \frac{\partial F}{\partial z}\,dx\,dy\,dz$$

Analogously, we get that

$$d(G\,dy\,dz) = \frac{\partial G}{\partial x}\,dx\,dy\,dz \quad \text{and} \quad d(H\,dz\,dx) = \frac{\partial H}{\partial y}\,dx\,dy\,dz$$

Therefore

$$d\eta = \left(\frac{\partial F}{\partial z} + \frac{\partial G}{\partial x} + \frac{\partial H}{\partial y}\right) dx\,dy\,dz$$

We have now developed all the concepts needed to reformulate Green's, Stokes', and Gauss' theorems in the language of forms.

THEOREM 13 (GREEN'S THEOREM). *Let $D$ be an elementary region in the $xy$-plane, with $\partial D$ given the counterclockwise orientation. Suppose $\omega = P(x, y)\,dx + Q(x, y)\,dy$ is a 1-form on some open set $K$ in $\mathbb{R}^3$ that contains $D$. Then*

$$\int_{\partial D} \omega = \int_D d\omega$$

Here $d\omega$ is a 2-form on $K$ and $D$ is in fact a surface in $\mathbb{R}^3$ parametrized by $\mathbf{\Phi}: D \to \mathbb{R}^3$, $\mathbf{\Phi}(x, y) = (x, y, 0)$. Since $P$ and $Q$ are explicitly *not* functions of $z$, then $\partial P/\partial z$ and $\partial Q/\partial z = 0$, and by Example 12, $d\omega = (\partial Q/\partial x - \partial P/\partial y)\,dx\,dy$. Consequently, Theorem 13 means nothing more than

$$\int_{\partial D} P\,dx + Q\,dy = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy$$

which is precisely Green's Theorem of Section 7.1. Hence Theorem 13 holds. Likewise, we have the following theorems.

THEOREM 14 (STOKES' THEOREM). *Let $S$ be an oriented surface in $\mathbb{R}^3$ with a boundary consisting of a simple closed surve $\partial S$ (Figure 7.6.3) oriented as the*



## OPTIONAL (*Continued*)

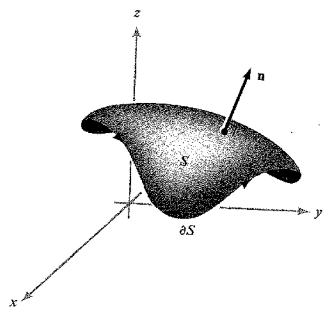



FIGURE 7.6.3
*An oriented surface to which Stokes' Theorem applies. ich the theory of forms applies.*

*boundary of $S$ (see Figure 7.2.1). Suppose that $\omega$ is a 1-form on some open set $K$ that contains $S$. Then*

$$\int_{\partial S} \omega = \int_S d\omega$$

THEOREM 15 (GAUSS' THEOREM). *Let $\Omega \subset \mathbb{R}^3$ be an elementary region with $\partial\Omega$ given the outward orientation (see Section 7.4). If $\eta$ is a 2-form on some region $K$ containing $\Omega$ then*

$$\int_{\partial\Omega} \eta = \int_\Omega d\eta$$

The reader has probably noticed the strong similarity in the statements of these theorems. In the vector field formulations, we used divergence for regions in $\mathbb{R}^3$ (Gauss' Theorem), and the curl for surfaces in $\mathbb{R}^3$ (Stokes' Theorem) and regions in $\mathbb{R}^2$ (Green's Theorem). Here we just use the unified notion of derivative of a differential form for all three theorems; and, in fact, we can state all theorems as one by introducing a little more terminology.

By an *oriented 2-manifold with boundary* in $\mathbb{R}^3$ we mean a surface in $\mathbb{R}^3$ whose boundary is a simple closed curve with orientation as described in Section 7.2. By an *oriented 3-manifold* in $\mathbb{R}^3$ we mean an elementary region

**OPTIONAL (*Continued*)**

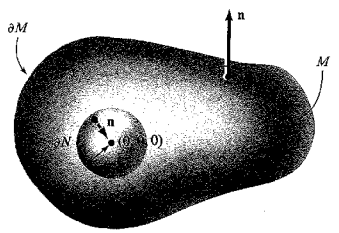



FIGURE 7.4.7
*Induced outward orientation on S.*

Now on $\partial N$, $\mathbf{n} = -\mathbf{r}/r$ and $r = \varepsilon$, since $\partial N$ is a sphere of radius $\varepsilon$, so

$$-\int_{\partial N} \frac{\mathbf{r}\cdot\mathbf{n}}{r^3}\,dS = \int_{\partial N} \frac{\varepsilon^2}{\varepsilon^4}\,dS = \frac{1}{\varepsilon^2}\int_{\partial N} dS$$

But $\int_{\partial N} dS = 4\pi\varepsilon^2$, the surface area of the sphere of radius $\varepsilon$. This proves the result.

Gauss' law has the following physical interpretation. The potential due to a point charge $Q$ at $(0, 0, 0)$ is given by

$$\phi(x, y, z) = \frac{Q}{4\pi r} = \frac{Q}{4\pi\sqrt{x^2 + y^2 + z^2}}$$

and the corresponding electric field is

$$\mathbf{E} = -\nabla\phi = \frac{Q}{4\pi}\left(\frac{\mathbf{r}}{r^3}\right)$$

Thus Theorem 10 states that the total electric flux $\int_{\partial M} \mathbf{E}\cdot d\mathbf{S}$ (that is, the flux of $\mathbf{E}$ out of a closed surface $\partial M$) equals $Q$ if the charge lies inside $M$ and zero otherwise. (A generalization is given in Exercise 10.) Note that even if $(0, 0, 0) \notin M$, $\mathbf{E}$ will still be nonzero on $M$.

For a continuous charge distribution described by a charge density $\rho$, the field $\mathbf{E}$ is related to the density $\rho$ by

$$\operatorname{div}\mathbf{E} = \nabla\cdot\mathbf{E} = \rho$$



Thus by Gauss' Theorem,

$$\int_S \mathbf{E}\cdot d\mathbf{S} = \int_\Omega \rho\,dV = Q$$

or the flux out of a surface is equal to the total charge inside.

**EXERCISES**

1. Let $S$ be a closed surface. Use Gauss' Theorem to show that if $\mathbf{F}$ is a $C^2$ vector field, then $\int_S (\nabla\times\mathbf{F})\cdot d\mathbf{S} = 0$. (Compare with Exercise 8 of Section 7.2.)
2. Let $\mathbf{F} = x^3\mathbf{i} + y^3\mathbf{j} + z^3\mathbf{k}$. Evaluate the surface integral of $\mathbf{F}$ over the unit sphere.
3. Evaluate $\int_{\partial\Omega} \mathbf{F}\cdot d\mathbf{S}$, where $\mathbf{F} = x\mathbf{i} + y\mathbf{j} - z\mathbf{k}$ and $\Omega$ is the unit cube (in the first octant). Perform the calculation directly and check by using the Divergence Theorem.
4. Repeat Exercise 3 for
   (a) $\mathbf{F} = \mathbf{i} + \mathbf{j} + \mathbf{k}$; and
   (b) $\mathbf{F} = x^2\mathbf{i} + y^2\mathbf{j} + z^2\mathbf{k}$.
5. Let $\mathbf{F} = y\mathbf{i} + z\mathbf{j} + xz\mathbf{k}$. Evaluate $\int_{\partial\Omega} \mathbf{F}\cdot d\mathbf{S}$ for each of the following regions $\Omega$.
   (a) $x^2 + y^2 \le z \le 1$
   (b) $x^2 + y^2 \le z \le 1$ and $x \ge 0$
   (c) $x^2 + y^2 \le z \le 1$ and $x \le 0$
6. Repeat Exercise 5 for $\mathbf{F} = (x - y)\mathbf{i} + (y - z)\mathbf{j} + (z - x)\mathbf{k}$.
7. Let $S$ be the surface of a region $\Omega$. Show that

$$\int_S \mathbf{r}\cdot\mathbf{n}\,dS = 3\ \text{volume}(\Omega)$$

   Attempt to explain this geometrically. (HINT: Assume $(0, 0, 0) \in \Omega$ and consider the skew cone with its vertex at $(0, 0, 0)$ with base $\Delta S$ and altitude $\|\mathbf{r}\|$. Its volume is $\frac{1}{3}(\Delta S)\cdot(\mathbf{r}\cdot\mathbf{n})$.)
8. Evaluate $\int_S \mathbf{F}\cdot\mathbf{n}\,dS$, where $\mathbf{F} = 3xy^2\mathbf{i} + 3x^2y\mathbf{j} + z^3\mathbf{k}$ and $S$ is the surface of the unit sphere.
9. Show $\int_\Omega (1/r^2)\,dx\,dy\,dz = \int_{\partial\Omega} (\mathbf{r}\cdot\mathbf{n}/r^2)\,dS$ where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.
10. Fix vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^3$ and numbers ("charges") $q_1, \ldots, q_k$. Set $\phi(x, y, z) = \sum_{i=1}^k q_i/(4\pi\|\mathbf{r} - \mathbf{v}_i\|)$, where $\mathbf{r} = (x, y, z)$. Show that for a closed surface $S$, and $\mathbf{E} = -\nabla\phi$,

$$\int_S \mathbf{E}\cdot d\mathbf{S} = Q$$

    where $Q$ is the total charge inside $S$. (Assume Gauss' law from Theorem 10 and that none of the charges are on $S$.)
11. Prove *Green's identities*

$$\int_{\partial\Omega} f\nabla g\cdot\mathbf{n}\,dS = \int_\Omega [f\nabla^2 g + \nabla f\cdot\nabla g]\,dV$$

$$\int_{\partial\Omega} (f\nabla g - g\nabla f)\cdot\mathbf{n}\,dS = \int_\Omega (f\nabla^2 g - g\nabla^2 f)\,dV$$

12. Suppose $\mathbf{F}$ satisfies $\operatorname{div}\mathbf{F} = 0$ and $\operatorname{curl}\mathbf{F} = \mathbf{0}$. Show that we can write $\mathbf{F} = \nabla f$, where $\nabla^2 f = 0$.

*13. Let $\rho$ be a continuous function on $\mathbb{R}^3$ such that $\rho(\mathbf{q}) = 0$ except for $\mathbf{q}$ in some region $\Omega$. Let $\mathbf{q} \in \Omega$ be denoted by $\mathbf{q} = (x, y, z)$. The *potential* of $\rho$ is defined as the function

$$\phi(\mathbf{p}) = \int_\Omega \frac{\rho(\mathbf{q})}{4\pi\|\mathbf{p} - \mathbf{q}\|}\, dV(\mathbf{q})$$

where $\|\mathbf{p} - \mathbf{q}\|$ is the distance between $\mathbf{p}$ and $\mathbf{q}$.

(a) Using the method of Theorem 10, show that $\int_{\partial W} \nabla\phi \cdot \mathbf{n}\, dS = -\int_W \rho\, dV$ for those regions $W$ that can be partitioned into a finite union of regions of type 4.

(b) Show that $\phi$ satisfies *Poisson's equation*

$$\nabla^2\phi = -\rho$$

(HINT: Use part (a)).

(Notice that if $\rho$ is a charge density, then the integral defining $\phi$ may be thought of as the sum of the potentials at $\mathbf{p}$ caused by point charges distributed over $\Omega$ according to the density $\rho$.)

14. Suppose $\mathbf{F}$ is tangent to the closed surface $S$ of a region $\Omega$. Prove that

$$\int_\Omega (\operatorname{div} \mathbf{F})\, dV = 0$$

*15. Use Gauss' law and symmetry to prove that the electric field due to a charge $Q$ evenly spread over the surface of a sphere is the same outside the surface as the field from a point charge $Q$ located at the center of the sphere. What is the field inside the sphere?

*16. Reformulate Exercise 15 in terms of gravitational fields.

17. Show how Gauss' law can be used to solve part (b) of Exercise 21 in Section 7.3.

*18. (*Transport Theorem*). Let $\phi(\mathbf{x}, t)$ be the flow of the vector field $\mathbf{F}$ on $\mathbb{R}^3$ (see Section 3.4), and let $J(\mathbf{x}, t)$ be the Jacobian of the map $\phi_t\colon \mathbf{x} \mapsto \phi(\mathbf{x}, t)$ for $t$ fixed.

(a) Using the proof of Theorem 3, Section 3.4, show that

$$\frac{\partial}{\partial t} J(\mathbf{x}, t) = [\operatorname{div} \mathbf{F}(\mathbf{x})]J(\mathbf{x}, t)$$

(b) Using the Change of Variables Theorem and (a), show that if $f(x, y, z, t)$ is a given function and $\Omega \subset \mathbb{R}^3$ is any region, then

$$\frac{d}{dt}\iiint_{\Omega_t} f(x, y, z, t)\, dx\, dy\, dz = \iiint_{\Omega_t} \left(\frac{Df}{Dt} + f \operatorname{div} \mathbf{F}\right) dx\, dy\, dz$$

(Transport equation)

where $\Omega_t = \phi_t(\Omega)$, which is the region moving with the flow, and where $Df/Dt = \partial f/\partial t + D_{\mathbf{x}} f \cdot \mathbf{F}$ is the material derivative (Exercise 7, Section 3.3).

(c) Taking $f = 1$ in (b), show that the following assertions are equivalent:

(*i*) $\operatorname{div} \mathbf{F} = 0$;

(*ii*) $\operatorname{volume}(\Omega_t) = \operatorname{volume}(\Omega)$; and

(*iii*) $J(\mathbf{x}, t) = 1$.

*19. Let $\phi, J, \mathbf{F}, f$ be as in Exercise 18. Prove the vector form of the Transport Theorem, namely,

$$\frac{d}{dt}\int_{\Omega_t} (f\mathbf{F})\, dx\, dy\, dz = \int_{\Omega_t} \left(\frac{\partial}{\partial t}(f\mathbf{F}) + \mathbf{F} \cdot \nabla(f\mathbf{F}) + (f\mathbf{F}) \operatorname{div} \mathbf{F}\right) dx\, dy\, dz$$



where $\mathbf{F} \cdot \nabla(f\mathbf{F})$ denotes the $3 \times 3$ derivative matrix $D(f\mathbf{F})$ operating on the column vector $\mathbf{F}$; in cartesian coordinates, $\mathbf{F} \cdot \nabla\mathbf{G}$ is the vector whose $i^{\text{th}}$ component is

$$\sum_{j=1}^{3} F_j \frac{\partial G^i}{\partial x_j}$$

(labeling $(x, y, z)$ as $(x_1, x_2, x_3)$).

**OPTIONAL**

## 7.5 APPLICATIONS TO PHYSICS AND DIFFERENTIAL EQUATIONS*

We can apply the concepts developed in this chapter to the formulation of some physical theories. Let us first discuss an important equation that is referred to as a *conservation* equation. For fluids, it expresses the conservation of mass, and for electromagnetic theory, the conservation of charge. We shall apply the equation to heat conduction and to electromagnetism.

Let $\mathbf{V}(t, x, y, z)$ be a $C^1$ vector field on $\mathbb{R}^3$ for each $t$ and let $\rho(t, x, y, z)$ be a $C^1$ real-valued function. By the *law of conservation of mass* for $\mathbf{V}$ and $\rho$, we shall mean that the condition

$$\frac{d}{dt}\int_\Omega \rho\, dV = -\int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n}\, dS$$

holds for all regions $\Omega$ in $\mathbb{R}^3$, where $\mathbf{J} = \rho\mathbf{V}$. See Figure 7.5.1.

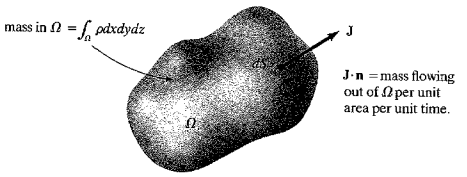



FIGURE 7.5.1

*The rate of change of mass in $\Omega$ equals the rate at which mass crosses $\partial\Omega$.*

* For additional examples the reader may profitably refer to H. M. Schey, *Div, Grad, Curl, and All That*, W. W. Norton, New York, 1973.

## OPTIONAL (*Continued*)

If we think of $\rho$ as a mass density ($\rho$ could also be charge density), that is, the mass per unit volume, and $\mathbf{V}$ as the velocity field of a fluid, the condition just says that the rate of change of total mass in $\Omega$ equals the rate at which mass flows *into* $\Omega$. Recall that $\int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n}\, dS$ is called the flux of $\mathbf{J}$. We need the following result.

THEOREM 11. *For* $\mathbf{V}$ *and* $\rho$ *defined on* $\mathbb{R}^3$, *the law of conservation of mass for* $\mathbf{V}$ *and* $\rho$ *is equivalent to the condition*

$$\operatorname{div} \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{1}$$

*that is*

$$\rho \operatorname{div} \mathbf{V} + \mathbf{V} \cdot \nabla \rho + \frac{\partial \rho}{\partial t} = 0 \tag{1'}$$

NOTE: Here, div $\mathbf{J}$ means that we compute div $\mathbf{J}$ for $t$ held fixed, and $\partial\rho/\partial t$ means we differentiate $\rho$ with respect to $t$ for $x$, $y$, $z$ fixed.

*Proof.* First, observe that $(d/dt) \int_\Omega \rho\, dx\, dy\, dz = \int_\Omega (\partial\rho/\partial t)\, dx\, dy\, dz$, and

$$\int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n}\, dS = \int_\Omega \operatorname{div} \mathbf{J}\, d\mathbf{V}$$

by the Divergence Theorem. Thus conservation of mass is equivalent to the condition

$$\int_\Omega \left( \operatorname{div} \mathbf{J} + \frac{\partial \rho}{\partial t} \right) dx\, dy\, dz = 0$$

Since this is to hold for all regions $\Omega$, this is equivalent to div $\mathbf{J} + \partial\rho/\partial t = 0$. ■

The equation div $\mathbf{J} + \partial\rho/\partial t = 0$ is called the *equation of continuity*. This is not the only equation governing fluid motion and it does not determine the motion of the fluid, but it is just one equation that must hold. Later in this section we shall obtain the additional equations needed to determine the flow. Using the Transport Theorem (Exercise 18, Section 7.4), the reader can check that the equation of continuity is equivalent to the condition

$$\frac{d}{dt} \int_{\Omega_t} \rho\, dx\, dy\, dz = 0$$



## OPTIONAL (*Continued*)

which says that the mass of a region moving with the fluid is constant in time. (The notation $\Omega_t$ is explained in the same exercise.)

The fluids that the continuity equation governs can be compressible. If div $\mathbf{V} = 0$ (incompressible case) and $\rho$ is constant, equation (1') follows automatically. But in general, even for incompressible fluids the equation is not automatic, because $\rho$ can depend on $(x, y, z)$ and $t$. Thus, while div $\mathbf{V} = 0$ may hold, $\operatorname{div}(\rho\mathbf{V}) \neq 0$ may still be true.

Next we discuss *Euler's equation for a perfect fluid*. Consider a homogeneous nonviscous fluid moving in space with a velocity field $\mathbf{V}$. The fluid being "perfect" means that if $\Omega$ is any portion of the fluid, forces of pressure act on the boundary of $\Omega$ along its normal. We assume that the force per unit area acting on $\partial\Omega$ is $-p\mathbf{n}$, where $p(x, y, z, t)$ is some function called the *pressure*. See Figure 7.5.2. Thus the total force acting on $\Omega$ is

$$\mathbf{F}_{\partial\Omega} = \text{Force} = -\int_{\partial\Omega} p\mathbf{n}\, dS$$

This is a *vector* quantity; the $i^{\text{th}}$ component of $\mathbf{F}_{\partial\Omega}$ is the integral of the $i^{\text{th}}$

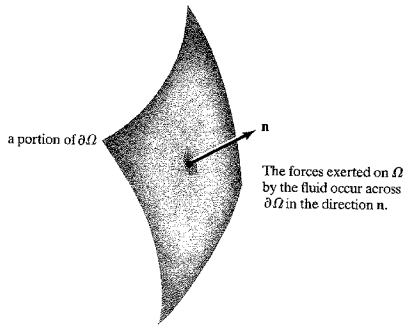



FIGURE 7.5.2
*The force acting on* $\partial\Omega$ *per unit area is* $-p\mathbf{n}$.

## OPTIONAL (*Continued*)

component of $p\mathbf{n}$ over the surface $\partial\Omega$ (this is therefore the surface integral of a real-valued function). If $\mathbf{e}$ is any fixed vector in space, we have

$$\mathbf{F}_{\partial\Omega} \cdot \mathbf{e} = -\int_{\partial\Omega} p\mathbf{e} \cdot \mathbf{n}\, dS$$

which is the integral of a scalar over $\partial\Omega$.

Notice that, by the divergence theorem and identity 8, p. 190,

$$\begin{aligned} \mathbf{e} \cdot \mathbf{F}_{\partial\Omega} &= -\int_{\Omega} \operatorname{div}(p\mathbf{e})\, dx\, dy\, dz \\ &= -\int_{\Omega} (\operatorname{grad} p) \cdot \mathbf{e}\, dx\, dy\, dz \end{aligned}$$

so

$$\mathbf{F}_{\partial\Omega} = -\int_{\Omega} \nabla p\, dx\, dy\, dz$$

Now we apply Newton's second law to a moving region $\Omega_t$. Here, $\Omega_t = \phi_t(\Omega)$, where $\phi_t(\mathbf{x}) = \phi(\mathbf{x}, t)$ denotes the flow of $\mathbf{V}$. The rate of change of momentum of $\Omega_t$ equals the force acting on it:

$$\frac{d}{dt}\int_{\Omega_t} \rho\mathbf{V}\, dx\, dy\, dz = \mathbf{F}_{\partial\Omega_t} = -\int_{\Omega_t} \nabla p\, dx\, dy\, dz$$

We apply the vector form of the Transport Theorem to the left-hand side (Exercise19, Section 7.4) to get

$$\int_{\Omega_t} \left\{ \frac{\partial}{\partial t}(\rho\mathbf{V}) + \mathbf{V} \cdot \nabla(\rho\mathbf{V}) + \rho\mathbf{V} \operatorname{div} \mathbf{V} \right\} dx\, dy\, dz = -\int_{\Omega_t} \nabla p\, dx\, dy\, dz$$

Since $\Omega_t$ is arbitrary, this is equivalent to

$$\frac{\partial}{\partial t}(\rho\mathbf{V}) + \mathbf{V} \cdot \nabla(\rho\mathbf{V}) + \rho\mathbf{V} \operatorname{div} \mathbf{V} = -\nabla p$$

Simplification using the equation of continuity (1′) gives

$$\rho\left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla\mathbf{V}\right) = -\nabla p \tag{2}$$

This is Euler's equation for a perfect fluid. For compressible fluids, $p$ is a given function of $\rho$ (for instance, for many gases, $p = A\rho^{\gamma}$ for constants $A$ and $\gamma$). If the fluid is, on the other hand, incompressible, $p$ is to be determined from the condition $\operatorname{div} \mathbf{V} = 0$. Equations (1) and (2) then govern the motion of the fluid completely. (An example that shows how complex particular solutions of (1) and (2) can be is given on the cover of this book.)



## HISTORICAL NOTE

The equations describing the motion of a fluid were first derived by Leonhard Euler in 1755, in a paper entitled "General Principles of the Motion of Fluids." Euler did basic work in mechanics as well as in pure mathematics; he essentially began the subject of analytical mechanics (as opposed to the geometric methods used by Newton). He is responsible for the equations of a rigid body (such as a tumbling satellite) and the formulation of many basic equations of mechanics as minimum principles. Euler wrote the first comprehensive textbook on calculus and contributed to virtually all branches of mathematics. He wrote several books and hundreds of research papers after he became totally blind, and he was working on a new treatise on fluid mechanics at the time of his death in 1783. Euler's equations for a fluid were eventually modified to include viscous effects by Navier and Stokes; the resulting Navier–Stokes' equations are described in virtually every textbook on fluid mechanics. Stokes is, of course, also responsible for Stokes' theorem, one of the main results in this book.

We shall now turn our attention to the *heat equation*, one of the most important equations of applied mathematics. It has been, and remains, one of the prime motivations for the study of partial differential equations.

Let us argue intuitively. If $T(t, x, y, z)$ (a $C^2$ function) denotes the temperature in a body at time $t$, then $\nabla T$ represents the temperature gradient and heat "flows" with the vector field $-\nabla T = \mathbf{F}$. Note that $\nabla T$ points in the direction of increasing $T$ (Chapter 3). Since heat flows from hot to cold, we have inserted a minus sign. The energy density, that is, the energy per unit volume, is $c\rho_0 T$, where $c$ is a constant (specific heat) and $\rho_0$ is the mass density, assumed constant. (We accept these assertions from elementary physics.) The *energy flux vector* is $\mathbf{J} = \kappa\mathbf{F}$, where $\kappa$ is a constant called the *conductivity*.

We propose that energy be conserved. Formally, this means that $\mathbf{J}$ and $\rho = c\rho_0 T$ should obey the law of conservation of mass with $\rho$ playing the role of "mass"; that is,

$$\frac{d}{dt}\int_{\Omega} \rho\, dV = -\int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n}\, dS$$

**OPTIONAL (*Continued*)**

By Theorem 11 this assertion is equivalent to

$$\operatorname{div} \mathbf{J} + \frac{\partial \rho}{\partial t} = 0$$

But $\operatorname{div} \mathbf{J} = \operatorname{div}(-\kappa \nabla T) = -\kappa \nabla^2 T$. (Recall that $\nabla^2 T = \partial^2 T/\partial x^2 + \partial^2 T/\partial y^2 + \partial^2 T/\partial z^2$ and $\nabla^2$ is the Laplace operator.) Continuing, we have $\partial \rho/\partial t = \partial(c\rho_0 T)/\partial t = c\rho_0(\partial T/\partial t)$. Thus equation (1) becomes, in this case,

$$\frac{\partial T}{\partial t} = \frac{\kappa}{c\rho_0} \nabla^2 T = k \nabla^2 T \tag{3}$$

where $k = \kappa/c\rho_0$ is called the *diffusivity*. Equation (3) is the important heat equation.

Just as (1) and (2) govern the flow of an ideal fluid, (3) governs the conduction of heat, in the following sense. If $T(0, x, y, z)$ is a given initial temperature distribution, then a unique $T(t, x, y, z)$ is determined that satisfies equation (3). In other words the initial condition, at $t = 0$, gives us the result for $t > 0$. Notice that if $T$ does not change with time (steady-state case) then we must have $\nabla^2 T = 0$ (Laplace's equation).

We shall now discuss *Maxwell's equations*, governing electromagnetic fields. The form of these equations depends on the physical units one is employing, and changing units introduces factors like $4\pi$, the velocity of light, and so on. We shall choose the system in which Maxwell's equations are simplest.

Let $\mathbf{E}(t, x, y, z)$ and $\mathbf{H}(t, x, y, z)$ be $C^1$ functions of $(t, x, y, z)$ that are vector fields for each $t$. They satisfy (by definition) Maxwell's equations with charge density $\rho(t, x, y, z)$ and current density $\mathbf{J}(t, x, y, z)$ when the following hold:

$$\nabla \cdot \mathbf{E} = \rho \qquad \text{(Gauss' law)} \tag{4}$$

$$\nabla \cdot \mathbf{H} = 0 \qquad \text{(no magnetic sources)} \tag{5}$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0} \qquad \text{(Faraday's law)} \tag{6}$$

and

$$\nabla \times \mathbf{H} - \frac{\partial \mathbf{E}}{\partial t} = \mathbf{J} \qquad \text{(Ampère's law)} \tag{7}$$

Of these laws, (4) and (6) were discussed earlier in Sections 7.4 and 7.2 in integral form; historically, they arose in these forms as physically observed



**OPTIONAL (*Continued*)**

laws. Ampère's law was mentioned for a special case in Example 12, Section 6.2.

Physically, one interprets $\mathbf{E}$ as the *electric field* and $\mathbf{H}$ as the *magnetic field*. As time $t$ progresses, these fields interact with each other and with any charges and currents that are present according to the above equations. For example, the propagation of electromagnetic waves in a vacuum is governed by these equations with $\mathbf{J} = \mathbf{0}$ and $\rho = 0$.

Since $\nabla \cdot \mathbf{H} = 0$ we can apply Theorem 8 of Section 7.3 to conclude that $\mathbf{H} = \nabla \times \mathbf{A}$ for some vector field $\mathbf{A}$. (We are assuming that $\mathbf{H}$ is defined on all of $\mathbb{R}^3$ for each time $t$.) This vector field $\mathbf{A}$ is not unique, and we can equally well use $\mathbf{A}' = \mathbf{A} + \nabla f$ for any function $f(t, x, y, z)$, since $\nabla \times \nabla f = \mathbf{0}$. (This freedom in the choice of $A$ is called *gauge freedom*.) For any such choice of $\mathbf{A}$, we have by (6)

$$\mathbf{0} = \nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \nabla \times \mathbf{E} + \frac{\partial}{\partial t} \nabla \times \mathbf{A}$$

$$= \nabla \times \mathbf{E} + \nabla \times \frac{\partial \mathbf{A}}{\partial t} = \nabla \times \left( \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right)$$

Hence applying Theorem 7, Section 7.3, there is a real-valued function $\phi$ on $\mathbb{R}^3$ such that

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \phi$$

Substituting this equation and $\mathbf{H} = \nabla \times \mathbf{A}$ into equation (7) and using the following identity from Table 3.1,

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$$

we get

$$\mathbf{J} = \nabla \times \mathbf{H} - \frac{\partial \mathbf{E}}{\partial t} = \nabla \times (\nabla \times \mathbf{A}) - \frac{\partial}{\partial t} \left( -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi \right)$$

$$= \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} + \frac{\partial^2 \mathbf{A}}{\partial t^2} + \frac{\partial}{\partial t} (\nabla \phi)$$

Thus

$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mathbf{J} + \nabla(\nabla \cdot \mathbf{A}) + \frac{\partial}{\partial t} (\nabla \phi)$$

that is

$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mathbf{J} + \nabla \left( \nabla \cdot \mathbf{A} + \frac{\partial \phi}{\partial t} \right) \tag{8}$$

**OPTIONAL (*Continued*)**

Again using the equation $\mathbf{E} + \partial\mathbf{A}/\partial t = -\nabla\phi$ and the equation $\nabla \cdot \mathbf{E} = \rho$, we obtain

$$\rho = \nabla \cdot \mathbf{E} = \nabla \cdot \left( -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \right) = -\nabla^2\phi - \frac{\partial(\nabla \cdot \mathbf{A})}{\partial t}$$

that is

$$\nabla^2\phi = -\rho - \frac{\partial(\nabla \cdot \mathbf{A})}{\partial t} \tag{9}$$

Now let us exploit the freedom in our choice of $\mathbf{A}$. We impose the "condition"

$$\nabla \cdot \mathbf{A} + \frac{\partial\phi}{\partial t} = 0 \tag{10}$$

We must be sure we can do this. Supposing we have a given $\mathbf{A}_0$ and a corresponding $\phi_0$, can we choose a new $\mathbf{A} = \mathbf{A}_0 + \nabla f$ and then a new $\phi$ such that $\nabla \cdot \mathbf{A} + \partial\phi/\partial t = 0$? With this new $\mathbf{A}$, the new $\phi$ is $\phi_0 - \partial f/\partial t$; we leave verification as an exercise for the reader. Condition (10) on $f$ then becomes

$$0 = \nabla \cdot (\mathbf{A}_0 + \nabla f) + \frac{\partial(\phi_0 - \partial f/\partial t)}{\partial t} = \nabla \cdot \mathbf{A}_0 + \nabla^2 f + \frac{\partial\phi_0}{\partial t} - \frac{\partial^2 f}{\partial t^2}$$

or

$$\nabla^2 f - \frac{\partial^2 f}{\partial t^2} = -\left( \nabla \cdot \mathbf{A}_0 + \frac{\partial\phi_0}{\partial t} \right) \tag{11}$$

Thus to be able to choose $\mathbf{A}$ and $\phi$ satisfying $\nabla \cdot \mathbf{A} + \partial\phi/\partial t = 0$, we must be able to solve equation (11) for $f$. One can indeed do this under general conditions, although we do not prove it here. Equation (11) is called the *inhomogeneous wave equation.*

If we accept that $\mathbf{A}$ and $\phi$ can be chosen to satisfy $\nabla \cdot \mathbf{A} + \partial\phi/\partial t = 0$, then the equations (8) and (9) for $\mathbf{A}$ and $\phi$ become

$$\nabla^2 \mathbf{A} - \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mathbf{J} \tag{8'}$$

$$\nabla^2\phi - \frac{\partial^2\phi}{\partial t^2} = -\rho \tag{9'}$$

Equation (9′) follows from (9) by substituting $-\partial\phi/\partial t$ for $\nabla \cdot \mathbf{A}$. Thus the wave equation appears again.



**OPTIONAL (*Continued*)**

Conversely, if $\mathbf{A}$ and $\phi$ satisfy the equations $\nabla \cdot \mathbf{A} + \partial\phi/\partial t = 0$, $\nabla^2\phi - \partial^2\phi/\partial t^2 = -\rho$, and $\nabla^2\mathbf{A} - \partial^2\mathbf{A}/\partial t^2 = -\mathbf{J}$, then $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$ and $\mathbf{H} = \nabla \times \mathbf{A}$ satisfy Maxwell's equations. *This procedure then reduces Maxwell's equations to a study of the wave equation.**

This is fortunate because the solutions to the wave equation have been well studied (one learns how to solve it in most courses in differential equations). To indicate the wavelike nature of the solutions, for example, observe that for any function $f$

$$\phi(t, x, y, z) = f(x - t)$$

solves the wave equation $\nabla^2\phi - (\partial^2\phi/\partial t^2) = 0$. This solution just propagates the graph of $f$ like a wave; thus one might conjecture that solutions of Maxwell's equations are wavelike in nature. Historically, this was Maxwell's great achievement, and it soon led to Hertz's discovery of radio waves.

Next we shall show briefly how vector analysis can be used to solve differential equations by a method called "potential theory" or "the Green's-function method." The presentation will be quite informal; the reader may consult the aforementioned references (see preceding footnote) for further information.

Suppose we wish to solve Poisson's equation $\nabla^2 u = \rho$ for $u(x, y, z)$ where $\rho(x, y, z)$ is a given function (this equation arises from Gauss' law if $\mathbf{E} = \nabla u$).

A function $G(\mathbf{x}, \mathbf{y})$ that has the properties

$$G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x}) \text{ and } \nabla^2 G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) \tag{12}$$

(in this expression $\mathbf{y}$ is held fixed), that is, which solves the differential equation with $\rho$ replaced by $\delta$, is called the *Green's function* for this differential equation. Here $\delta(\mathbf{x} - \mathbf{y})$ represents the Dirac delta function, "defined" by†

(*i*) $\delta(\mathbf{x} - \mathbf{y}) = 0$ for $\mathbf{x} \neq \mathbf{y}$ and

(*ii*) $\int_{\mathbb{R}^3} \delta(\mathbf{x} - \mathbf{y})\, d\mathbf{y} = 1$.

* There are variations on this procedure. For further details see, for example, G. F. D. Duff and D. Naylor, *Differential Equations of Applied Mathematics*, Wiley, New York, 1966, or books on electromagnetic theory, such as J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, 1962.

† This is not a precise definition; nevertheless, it is enough here to assume that $\delta$ is a symbolic expression with the operational property (13). See the references in the preceding footnote for a more careful definition of $\delta$.

**OPTIONAL (*Continued*)**

It has the following operational property that formally follows from (*i*) and (*ii*): For any continuous function $f(\mathbf{x})$

$$\int_{\mathbb{R}^3} f(\mathbf{y})\delta(\mathbf{x} - \mathbf{y})\, d\mathbf{y} = f(\mathbf{x}) \tag{13}$$

This is sometimes called the *sifting property* of $\delta$.

THEOREM 12. *If $G(\mathbf{x}, \mathbf{y})$ satisfies the differential equation $\nabla^2 u = \rho$ with $\rho$ replaced by $\delta(\mathbf{x} - \mathbf{y})$, then*

$$u(\mathbf{x}) = \int_{\mathbb{R}^3} G(\mathbf{x}, \mathbf{y})\rho(\mathbf{y})\, d\mathbf{y} \tag{14}$$

*is a solution to $\nabla^2 u = \rho$.*

*Proof.* To see this, note that

$$\begin{aligned} \nabla^2 \int_{\mathbb{R}^3} G(\mathbf{x}, \mathbf{y})\rho(\mathbf{y})\, d\mathbf{y} &= \int_{\mathbb{R}^3} (\nabla^2 G(\mathbf{x}, \mathbf{y}))\rho(\mathbf{y})\, d\mathbf{y} \\ &= \int_{\mathbb{R}^3} \delta(\mathbf{x} - \mathbf{y})\rho(\mathbf{y})\, d\mathbf{y} && \text{by (12)} \\ &= \rho(\mathbf{x}) && \text{by (13)} \ \blacksquare \end{aligned}$$

The "function" $\rho(\mathbf{x}) = \delta(\mathbf{x})$ represents a unit charge concentrated at a single point (see conditions (*i*) and (*ii*) above). Thus $G(\mathbf{x}, \mathbf{y})$ *represents the potential at* $\mathbf{x}$ *due to a charge placed at* $\mathbf{y}$.

We claim that equation (12) is satisfied if we choose

$$G(\mathbf{x}, \mathbf{y}) = -\frac{1}{4\pi\|\mathbf{x} - \mathbf{y}\|} \tag{15}$$

Clearly $G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x})$. To check the second part of (12), we must verify that $\nabla^2 G(\mathbf{x}, \mathbf{y})$ has the following two formal properties of the $\delta$-function:

(*i*) $\nabla^2 G(\mathbf{x}, \mathbf{y}) = 0$ for $\mathbf{x} \neq \mathbf{y}$; and

(*ii*) $\int_{\mathbb{R}^3} \nabla^2 G(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = 1$

Property (*i*) is true because the gradient of $G$ is

$$\nabla G(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{r}}{4\pi r^3}$$

where $\mathbf{r} = \mathbf{x} - \mathbf{y}$ is the vector from $\mathbf{y}$ to $\mathbf{x}$ and $r = \|\mathbf{r}\|$ (see Exercise 8, Section 3.5), and therefore for $r \neq 0$, $\nabla \cdot \nabla G(\mathbf{x}, \mathbf{y}) = 0$ (as in the aforementioned exercise). For (*ii*) let $B$ be a ball about $\mathbf{x}$; by (*i*),



**OPTIONAL (*Continued*)**

$$\int_{\mathbb{R}^3} \nabla^2 G(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = \int_B \nabla^2 G(\mathbf{x}, \mathbf{y})\, d\mathbf{y}$$

This, in turn, equals

$$\int_{\partial B} \nabla G(\mathbf{x}, \mathbf{y}) \cdot \mathbf{n}\, dS$$

by Gauss' Theorem. Thus, by Theorem 10,

$$\int_{\partial B} \nabla G(\mathbf{x}, \mathbf{y}) \cdot \mathbf{n}\, dS = \int_{\partial B} \frac{\mathbf{r} \cdot \mathbf{n}}{4\pi r^3}\, dS = 1$$

which proves (*ii*).

Thus, *the solution of* $\nabla^2 u = \rho$ *is*

$$u(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{-\rho(\mathbf{y})}{4\pi\|\mathbf{x} - \mathbf{y}\|}\, d\mathbf{y} \tag{16}$$

by Theorem 12.

In two dimensions, one can similarly show that

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{2\pi} \log\|\mathbf{x} - \mathbf{y}\| \tag{17}$$

so the solution of $\nabla^2 u = \rho$ is

$$u(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \rho(\mathbf{y}) \log\|\mathbf{x} - \mathbf{y}\|\, d\mathbf{y} \tag{18}$$

We now turn to the problem of using Green's functions to solve Poisson's equation in a bounded region with given boundary conditions. To do this, we need Green's first and second identities, which can be obtained from the divergence theorem. We start with the identity

$$\int_V \nabla \cdot \mathbf{F}\, dV = \int_S \mathbf{F} \cdot \mathbf{n}\, dS$$

where $V$ is a region in space, $S$ is its boundary, and $\mathbf{n}$ is the outward unit normal vector at any point on $S$. Replacing $\mathbf{F}$ by $f\nabla g$, where $f$ and $g$ are scalar functions, we obtain

$$\int_V \nabla f \cdot \nabla g\, dV + \int_V f\nabla^2 g\, dV = \int_S f \frac{\partial g}{\partial n}\, dS \tag{19}$$

**OPTIONAL (*Continued*)**

where $\partial g/\partial n = \nabla g \cdot \mathbf{n}$. This is *Green's first identity*. If we simply permute $f$ and $g$, and subtract the result from the above equation, we obtain *Green's second identity*

$$\int_V (f\nabla^2 g - g\nabla^2 f)\, dV = \int_S \left(f\frac{\partial g}{\partial n} - g\frac{\partial f}{\partial n}\right) dS \tag{20}$$

It is this identity that we shall use.

Consider Poisson's equation

$$\nabla^2 u = \rho$$

in some region $V$, and the corresponding equations for the Green's function

$$G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x}) \text{ and } \nabla^2 G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$$

Inserting $u$ and $G$ into (20) we obtain

$$\int_V (u\nabla^2 G - G\nabla^2 u)\, dV = \int_S \left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right) dS$$

Choosing our integration variable to be $\mathbf{y}$ and using $G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x})$, this becomes

$$\int_V [u(\mathbf{y})\delta(\mathbf{x} - \mathbf{y}) - G(\mathbf{x}, \mathbf{y})\rho(\mathbf{y})]\, d\mathbf{y} = \int_S \left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right) dS$$

that is

$$u(\mathbf{x}) = \int_V G(\mathbf{x}, \mathbf{y})\rho(\mathbf{y})\, d\mathbf{y} + \int_S \left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right) dS \tag{21}$$

Note that for an unbounded region, this becomes identical to our previous result (14) for all of space. Equation (21) enables us to solve for $u$ in a bounded region where $\rho = 0$ by incorporating the conditions that $u$ must obey on $S$.

If $\rho = 0$, (21) reduces to

$$u = \int_S \left(u\frac{\partial G}{\partial n} - G\frac{\partial u}{\partial n}\right) dS$$

or fully

$$u(\mathbf{x}) = \int_S \left[u(\mathbf{y})\frac{\partial G}{\partial n}(\mathbf{x}, \mathbf{y}) - G(\mathbf{x}, \mathbf{y})\frac{\partial u}{\partial n}(\mathbf{y})\right] dS(\mathbf{y}) \tag{22}$$

where $u$ appears on both sides of the equation. The crucial point is that



**OPTIONAL (*Continued*)**

evaluation of the integral requires only that we know the behavior of $u$ on $S$. Commonly either $u$ is given on the boundary (*Dirichlet problem*) or $\partial u/\partial n$ is given on the boundary (*Neumann problem*). If we know $u$ on the boundary, we want to make $G\ \partial u/\partial n$ vanish on the boundary so we can evaluate the integral. So if $u$ is given on $S$ we must find a $G$ such that $G(\mathbf{x}, \mathbf{y})$ vanishes whenever $\mathbf{y}$ lies on $S$. This is called the *Dirichlet Green's function for the region V*. Conversely, if $\partial u/\partial n$ is given on $S$ we must find a $G$ such that $\partial G/\partial n$ vanishes on $S$. This is the *Neumann Green's function*.

Thus, a Dirichlet Green's function $G(\mathbf{x}, \mathbf{y})$ is defined for $\mathbf{x}$ and $\mathbf{y}$ in the volume $V$ and satisfies these three conditions:

(a) $G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x})$
(b) $\nabla^2 G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$ and
(c) $G(\mathbf{x}, \mathbf{y}) = 0$ when $\mathbf{y}$ lies on $S$, the boundary of the region $V$

(Note that by (a), in (b) and (c) the variables $\mathbf{x}$ and $\mathbf{y}$ can be interchanged without changing the condition).

It is perhaps surprising that *condition* (a) *is actually a consequence of* (b) *and* (c), *provided* (b) *and* (c) *also hold with* $\mathbf{x}$ *and* $\mathbf{y}$ *interchanged.*

To see this, we fix $\mathbf{y}$ and use (20) with $f(\mathbf{x}) = G(\mathbf{x}, \mathbf{y})$ and $g(\mathbf{x}) = G(\mathbf{y}, \mathbf{x})$. By (b), $\nabla^2 f(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$ and $\nabla^2 g(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{y})$, and by (c), $f$ and $g$ vanish on $S$. Thus, (20) becomes

$$\int_V f(\mathbf{x})\delta(\mathbf{x} - \mathbf{y}) - g(\mathbf{x})\delta(\mathbf{x} - \mathbf{y})\, dV = 0$$

or

$$f(\mathbf{x}) - g(\mathbf{x}) = 0.$$

Thus $G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x})$. This means, in effect, that in examples it is not necessary to check condition (a). (This result is sometimes called the *principle of reciprocity.*)

Doing any particular Dirichlet or Neumann problem thus becomes the task of finding the appropriate Green's function. We shall do this by modifying the Green's function for Laplace's equations on all $\mathbb{R}^2$ or $\mathbb{R}^3$, namely (16) and (17).

As an example, we shall now use the two-dimensional Green's-function method to construct the Dirichlet Green's function for the disc of radius $R$ (see Figure 7.5.3). This will enable us to solve $\nabla^2 u = 0$ (or $\nabla^2 u = \rho$) with $u$ given on the boundary circle.

**OPTIONAL (*Continued*)**

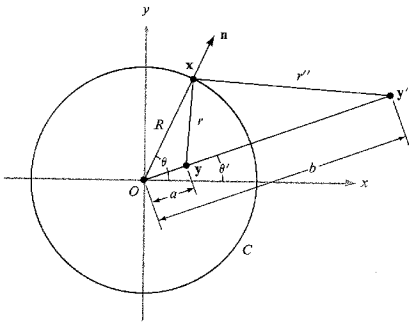



FIGURE 7.5.3
*Geometry of the construction of the Green's function for a disc.*

In Figure 7.5.3 we have drawn the point $\mathbf{x}$ on the circumference because that is where we want $G$ to vanish.* The Green's function $G(\mathbf{x}, \mathbf{y})$ that we shall find will, of course, be valid for all $\mathbf{x}$, $\mathbf{y}$ in the disc. The point $\mathbf{y}'$ represents the "reflection" of the point $\mathbf{y}$ into the region outside the circle, such that $ab = R^2$. Now when $\mathbf{x} \in C$, by the similarity of the triangles $\mathbf{x}O\mathbf{y}$ and $\mathbf{x}O\mathbf{y}'$,

$$\frac{r}{R} = \frac{r''}{b}$$

or

$$r = \frac{r''R}{b} = \frac{r''a}{R}$$

So if we choose our Green's function to be

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{2\pi}\left(\log r - \log \frac{r''a}{R}\right) \tag{23}$$

we see that $G$ is zero if $\mathbf{x}$ is on $C$. Since $r''a/R$ reduces to $r$ when $\mathbf{y}$ is on $C$, $G$ also vanishes when $\mathbf{y}$ is on $C$. If we can show that $G$ satisfies $\nabla^2 G = \delta(\mathbf{x} - \mathbf{y})$

* According to the procedure above, $G(\mathbf{x}, \mathbf{y})$ is supposed to vanish when either $\mathbf{x}$ or $\mathbf{y}$ is on $C$. We have chosen $\mathbf{x}$ on $C$ to begin with.



**OPTIONAL (*Continued*)**

in the circle, then we will have proved that $G$ is indeed the Dirichlet Green's function. From (17) we know that $\nabla^2(\log r)/2\pi = \delta(\mathbf{x} - \mathbf{y})$, so

$$\nabla^2 G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y}) - \delta(\mathbf{x} - \mathbf{y}')$$

but $\mathbf{y}'$ is always outside the circle, so $\mathbf{x}$ can never be equal to $\mathbf{y}'$ and so $\delta(\mathbf{x} - \mathbf{y}')$ is always zero. Hence

$$\nabla^2 G(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})$$

and thus $G$ is the Dirichlet Green's function for the circle.

Now we shall consider the problem of solving

$$\nabla^2 u = 0$$

in this circle if $u(R, \theta) = f(\theta)$ is the given boundary condition. By (22) we have a solution

$$u = \int_C \left(u \frac{\partial G}{\partial n} - G \frac{\partial u}{\partial n}\right) dS$$

But $G = 0$ on $C$, so we are left with the integral

$$u = \int_C u \frac{\partial G}{\partial n}\, dS$$

where we can replace $u$ by $f(\theta)$ since the integral is around $C$. Thus the task of solving the Dirichlet problem in the circle is reduced to finding $\partial G/\partial n$. From (23) we can write

$$\frac{\partial G}{\partial n} = \frac{1}{2\pi}\left(\frac{1}{r}\frac{\partial r}{\partial n} - \frac{1}{r''}\frac{\partial r''}{\partial n}\right)$$

Now

$$\frac{\partial r}{\partial n} = \nabla r \cdot \mathbf{n}$$

and

$$\nabla r = \frac{\mathbf{r}}{r}$$

where $\mathbf{r} = \mathbf{x} - \mathbf{y}$, so

$$\frac{\partial r}{\partial n} = \frac{\mathbf{r} \cdot \mathbf{n}}{r} = \frac{r\cos(nr)}{r} = \cos(nr)$$

where $(nr)$ represents the angle between $\mathbf{n}$ and $\mathbf{r}$. Likewise

$$\frac{\partial r''}{\partial n} = \cos(nr'')$$

**OPTIONAL (*Continued*)**

Now, in triangle $\mathbf{xy}O$, we have, by the cosine law

$$a^2 = r^2 + R^2 - 2rR\cos(nr)$$

and in triangle $\mathbf{xy}'O$, we get

$$b^2 = (r'')^2 + R^2 - 2r''R\cos(nr'')$$

and so

$$\frac{\partial r}{\partial n} = \cos(nr) = \frac{R^2 + r^2 - a^2}{2rR}$$

and

$$\frac{\partial r''}{\partial n} = \cos(nr'') = \frac{R^2 + (r'')^2 - b^2}{2r''R}$$

Hence

$$\frac{\partial G}{\partial n} = \frac{1}{2\pi}\left[\frac{R^2 + r^2 - a^2}{2r^2R} - \frac{R^2 + (r'')^2 - b^2}{2(r'')^2R}\right]$$

Using the relationship between $r$ and $r''$ when $\mathbf{x}$ is on $C$, we get

$$\left.\frac{\partial G}{\partial n}\right|_{\mathbf{x}\in C} = \frac{1}{2\pi}\left(\frac{R^2 - a^2}{Rr^2}\right)$$

Thus the solution can be written as

$$u = \frac{1}{2\pi}\int_C f(\theta)\frac{R^2 - a^2}{Rr^2}\,ds$$

Let us write this in a more explicit and tractable form. First, note that in triangle $\mathbf{xy}O$, we can write

$$r = [a^2 + R^2 - 2aR\cos(\theta - \theta')]^{1/2}$$

where $\theta$ and $\theta'$ are the polar angles in $\mathbf{x}$- and $\mathbf{y}$-space, respectively. Second, our solution must be valid for all $\mathbf{y}$ in the circle; hence the distance of $\mathbf{y}$ from the origin must now become a variable, which we shall call $r'$. Finally, note that $ds = R\,d\theta$ on $C$, so we can write the solution in polar coordinates as

$$u(r', \theta') = \frac{R^2 - (r')^2}{2\pi}\int_0^{2\pi}\frac{f(\theta)\,d\theta}{(r')^2 + R^2 - 2r'R\cos(\theta - \theta')}$$

This is known as *Poisson's formula in two dimensions.** As an exercise, the reader should use this to write down the solution of $\nabla^2 u = \rho$ with $u$ a given function $f(\theta)$ on the boundary.

* There are several ways of deriving this famous formula. For the method of complex variables, see J. Marsden, *Basic Complex Analysis*, W. H. Freeman and Company, San Francisco, 1973, p. 145. For the method of Fourier series, see J. Marsden, *Elementary Classical Analysis*, W. H. Freeman and Company, 1974, p. 466.



**OPTIONAL (*Continued*)**

## EXERCISES

1. (a) Supply the details for the assertion on p. 452 that

$$\frac{d}{dt}\int_{\Omega_t}\rho\,dx\,dy\,dz = 0$$

is equivalent to the law of conservation of mass.

(b) Using (a) and the Change of Variables Theorem show that $\rho(\mathbf{x}, t)$ can be expressed in terms of the Jacobian $J(\mathbf{x}, t)$ of the flow map $\phi(\mathbf{x}, t)$ and $\rho(\mathbf{x}, 0)$ by the equation

$$\rho(\mathbf{x}, t)J(\mathbf{x}, t) = \rho(\mathbf{x}, 0)$$

(c) What can you conclude from (b) for incompressible flow?

2. Let $\mathbf{V}$ be a vector field with flow $\phi(\mathbf{x}, t)$ and let $\mathbf{V}$ and $\rho$ satisfy the law of conservation of mass. Let $\Omega_t$ be the region transported with the flow. Prove the following version of the Transport Theorem (see Exercise 18, Section 7.4):

$$\frac{d}{dt}\int_{\Omega_t}\rho f\,dx\,dy\,dz = \int_{\Omega_t}\rho\frac{Df}{Dt}\,dx\,dy\,dz$$

3. (*Bernoulli's law*) (a) Let $\mathbf{V}$, $\rho$ satisfy the law of conservation of mass and equation (2) (Euler's equation for a perfect fluid). Suppose $\mathbf{V}$ is irrotational and hence that $\mathbf{V} = \mathbf{V}\phi$ for a function $\phi$. Show that if $C$ is a path connecting two points $P_1$ and $P_2$, then

$$\left(\frac{\partial\phi}{\partial t} + \frac{1}{2}\|\mathbf{V}\|^2\right)\Big|_{P_1}^{P_2} + \int_C\frac{dp}{\rho} = 0$$

(HINT: You will require the identity

$$\mathbf{V}\cdot(\nabla\mathbf{V}) = \tfrac{1}{2}\nabla(\|\mathbf{V}\|^2) + (\nabla\times\mathbf{V})\times\mathbf{V}$$

from Table 3.1, p. 190.)

(b) If in (a), $\mathbf{V}$ is stationary—that is, $\partial\mathbf{V}/\partial t = 0$—and $\rho$ is constant, show that

$$\frac{1}{2}\|\mathbf{V}\|^2 + \frac{p}{\rho}$$

is constant in space. Deduce that, in this situation, "higher pressure is associated with lower fluid speed."

4. Using Exercise 3, show that if $\phi$ satisfies Laplace's equation $\nabla^2\phi = 0$, then $\mathbf{V} = \nabla\phi$ is a stationary solution to Euler's equation for a perfect *incompressible* fluid.

5. Verify that Maxwell's equations imply the equation of continuity for $\mathbf{J}$ and $\rho$.

6. Let $H$ denote the upper half space $z \geq 0$. For a point $\mathbf{x} = (x, y, z)$ in $H$, let $R(\mathbf{x}) = (x, y, -z)$, the reflection of $\mathbf{x}$ in the $xy$-plane. Let $G(\mathbf{x}, \mathbf{y}) = -1/4\pi\|\mathbf{x} - \mathbf{y}\|$ be the Green's function for all of $\mathbb{R}^3$.

(a) Verify that the function $\tilde{G}$ defined by

$$\tilde{G}(\mathbf{x}, \mathbf{y}) = G(\mathbf{x}, \mathbf{y}) - G(R(\mathbf{x}), \mathbf{y})$$

is the Green's function for the Laplacian in $H$.

**OPTIONAL (*Continued*)**

Now

$$dx \wedge (dx \wedge dy) = (dx \wedge dx) \wedge dy = 0 \wedge dy = 0$$
$$dy \wedge (dx \wedge dy) = -dy \wedge (dy \wedge dx)$$
$$= -(dy \wedge dy) \wedge dx = 0 \wedge dx = 0$$

and

$$dz \wedge (dx \wedge dy) = (-1)^2(dx \wedge dy) \wedge dz = dx\,dy\,dz$$

Consequently

$$d(F\,dx\,dy) = \frac{\partial F}{\partial z}\,dx\,dy\,dz$$

Analogously, we get that

$$d(G\,dy\,dz) = \frac{\partial G}{\partial x}\,dx\,dy\,dz \quad \text{and} \quad d(H\,dz\,dx) = \frac{\partial H}{\partial y}\,dx\,dy\,dz$$

Therefore

$$d\eta = \left(\frac{\partial F}{\partial z} + \frac{\partial G}{\partial x} + \frac{\partial H}{\partial y}\right) dx\,dy\,dz$$

We have now developed all the concepts needed to reformulate Green's, Stokes', and Gauss' theorems in the language of forms.

THEOREM 13 (GREEN'S THEOREM). *Let $D$ be an elementary region in the $xy$-plane, with $\partial D$ given the counterclockwise orientation. Suppose $\omega = P(x, y)\,dx + Q(x, y)\,dy$ is a 1-form on some open set $K$ in $\mathbb{R}^3$ that contains $D$. Then*

$$\int_{\partial D} \omega = \int_D d\omega$$

Here $d\omega$ is a 2-form on $K$ and $D$ is in fact a surface in $\mathbb{R}^3$ parametrized by $\mathbf{\Phi}: D \to \mathbb{R}^3$, $\mathbf{\Phi}(x, y) = (x, y, 0)$. Since $P$ and $Q$ are explicitly *not* functions of $z$, then $\partial P/\partial z$ and $\partial Q/\partial z = 0$, and by Example 12, $d\omega = (\partial Q/\partial x - \partial P/\partial y)\,dx\,dy$. Consequently, Theorem 13 means nothing more than

$$\int_{\partial D} P\,dx + Q\,dy = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy$$

which is precisely Green's Theorem of Section 7.1. Hence Theorem 13 holds. Likewise, we have the following theorems.

THEOREM 14 (STOKES' THEOREM). *Let $S$ be an oriented surface in $\mathbb{R}^3$ with a boundary consisting of a simple closed surve $\partial S$ (Figure 7.6.3) oriented as the*



**OPTIONAL (*Continued*)**

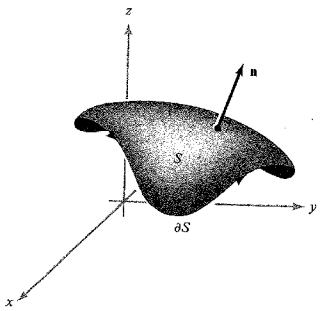



FIGURE 7.6.3
*An oriented surface to which Stokes' Theorem applies.*
*ich the theory of forms applies.*

*boundary of $S$ (see Figure 7.2.1). Suppose that $\omega$ is a 1-form on some open set $K$ that contains $S$. Then*

$$\int_{\partial S} \omega = \int_S d\omega$$

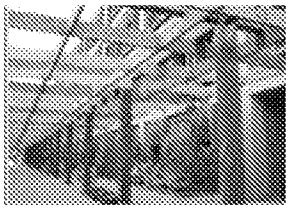
THEOREM 15 (GAUSS' THEOREM). *Let $\Omega \subset \mathbb{R}^3$ be an elementary region with $\partial\Omega$ given the outward orientation (see Section 7.4). If $\eta$ is a 2-form on some region $K$ containing $\Omega$ then*

$$\int_{\partial\Omega} \eta = \int_\Omega d\eta$$

The reader has probably noticed the strong similarity in the statements of these theorems. In the vector field formulations, we used divergence for regions in $\mathbb{R}^3$ (Gauss' Theorem), and the curl for surfaces in $\mathbb{R}^3$ (Stokes' Theorem) and regions in $\mathbb{R}^2$ (Green's Theorem). Here we just use the unified notion of derivative of a differential form for all three theorems; and, in fact, we can state all theorems as one by introducing a little more terminology.

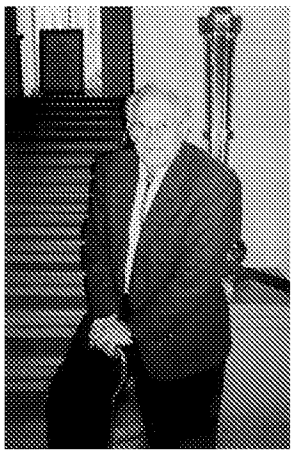
By an *oriented 2-manifold with boundary* in $\mathbb{R}^3$ we mean a surface in $\mathbb{R}^3$ whose boundary is a simple closed curve with orientation as described in Section 7.2. By an *oriented 3-manifold* in $\mathbb{R}^3$ we mean an elementary region

# APPENDIX N



# Lawrence W. Stark dies at 78

November 1, 2004



**Related links:**
UC press release
Telerobotics & neurology
Conversations with history

Return to Announcements

Dr. Lawrence W. Stark, a professor emeritus of physiological optics and engineering at the University of California, Berkeley, recognized worldwide as a pioneer in the use of control and information theory to characterize neurological systems, died Friday, Oct. 22.

Stark died of cancer at his home in Berkeley at the age of 78, four years after being diagnosed with Non-Hodgkin's lymphoma.

"Stark's death is a huge loss for the world of optometry," said Dennis Levi, dean of UC Berkeley's School of Optometry. "His landmark studies on eye movement control truly advanced the field of vision science."

A neurologist by training, Stark is credited for seminal research that applied engineering principles, particularly control theory, to biological systems.

"Stark was unique in his ability to identify aspects of engineering analysis relevant to medicine and biology," said Laurence Young, Apollo Program professor of astronautics at the Massachusetts Institute of Technology and Stark's first graduate student at MIT. "A perfect example is the seminal work he conducted in analyzing the way a pupil reacts to light in terms of a linear control system."

Young pointed out that those same principles have been applied to such areas as pilot control of airplanes and spacecraft. "Characterizing how the eye moves and how the brain processes visual cues is essential to understanding how pilots control airplanes, and why people get motion sickness," he said.

In addition to analyzing the feedback control system governing pupil contractions, Stark also developed the scan path theory of eye movements. He studied the way people's brains viewed the world and analyzed the vast number of jumping eye movements, or "saccades," people make. He noticed specific sequences to how people glimpsed a room, face or other scene before them, and realized how those sequences provided clues to the importance of pictures generated by the brain.

Stark's drive to understand visual processes within an engineering discipline led to his later research interests in robotic vision and virtual reality.

"Although he was trained in medicine, he was very interested in the physical sciences and engineering," said Gerald Westheimer, UC Berkeley professor of neurobiology. "He was a true crossover scientist, bringing applied engineering concepts to neurological functions, and the variables inherent in biology to engineering."

Stark was born in New York on Feb. 21, 1926. He credited his early interest in engineering to the influence of his father, an MIT-educated chemical engineer. He once recalled in an interview how, as a young boy, he took apart his mother's typewriter and put it back together again — minus four screws. His mother was impressed with his success, he said, until the typewriter stopped working a few weeks later.

Undaunted, Stark maintained his curiosity for how things work, going on to Columbia University and majoring in English, biology and zoology. After receiving his bachelor's degree in 1945, he joined the military, taking the U.S. Navy up on its offer to pay for his medical school tuition. Stark went to New York's Albany Medical College, where he earned his M.D. in 1948.

He then spent two years in England at Oxford University and University College, conducting research in neurophysiology, biochemistry and biophysics, before returning to the U.S. Navy to serve as a doctor during the Korean War.

In 1954, after the war ended, Stark joined Yale University as an assistant professor of medicine.

In 1960, he became head of the neurology section of MIT's Center for Communication Sciences, and in 1965, he founded and became chairman of the Biomedical Engineering Department, one of the country's first bioengineering departments, at the University of Illinois at Chicago.

"He was really one of the first people ever to use engineering theory to study a physiological system," said Blake Hannaford, director of the Biorobotics Laboratory at the University of Washington and one of Stark's former Ph.D. students. "He played a pivotal role in the 1950s and 1960s in founding the field of bioengineering."

Stark joined the UC Berkeley faculty in 1968 with joint appointments as professor at the School of Optometry, the Department of Electrical Engineering and Computer Sciences and the Department of Mechanical Engineering. He also collaborated on neuro-ophthalmology research with colleagues at UC San Francisco.

He retired from UC Berkeley in 1994, but remained active in his lab on campus.

"He was always questioning, bringing up new ideas," said Elwin Marg, UC Berkeley professor emeritus of vision science and optometry and a close colleague of Stark's for nearly 50 years. "His enthusiasm and curiosity were an inspiration to his students."

So many of his students went on to distinguish themselves in academic careers that a scientific conference was held in his honor. In 1994, John Semmlow, a professor of bioengineering at Rutgers University in New Jersey and one of Stark's former students, organized the First International Conference on Vision and Movement in Man and Machine, a two-day symposium attended primarily by Stark's colleagues and former students and affectionately nicknamed "Starkfest."

Although research papers were presented at the conference and later published in peer-reviewed journals, including special issues of the Annals of Biomedical Engineering and Optometry and Vision Science, participants viewed the event as a chance to honor their former advisor and renew old friendships, said Semmlow.

"Stark maintained a strong bond with nearly all of his students," said Semmlow, one of two students from the University of Illinois who followed Stark to UC Berkeley. "He had an extremely dynamic personality. He was also extraordinarily intelligent, very well read, and he cast his interests in many different directions."

Two more Starkfest meetings have been held since 1994, most recently in Marseilles, France, in 2002.

Stark received numerous honors throughout his career, including an honorary Sc.D. from the State University of New York and an honorary Ph.D. from Tokushima University in Japan. He was named a Guggenheim Fellow in 1968, a fellow of the Institute of Electrical and Electronics Engineers in 1970, a recipient of the William J. Morlock Award in Biomedical

Engineering in 1977, and a fellow of the American Institute of Medical and Biological Engineering in 1992.

Stark is survived by his partner of 18 years, Jill Strohn of Berkeley, Calif.; three daughters, Stefanie Stark of Kensington, Calif., Nanou Matteson of Berkeley, Calif., and Elizabeth Stark of San Francisco; Elizabeth's mother, Wendy Bartlett of Berkeley, Calif.; ex-wife, Jeanne Stark-Iochmans of Berkeley, Calif.; his brother, Matthew of Minneapolis, Minn.; and four grandchildren.

Stark's first wife and Stefanie's mother, Jane Stark, died in 2001.

Memorial services at the UC Berkeley Faculty Club are being planned. When the exact date and time are available, they will be posted online at http://scan.berkeley.edu/larry. In lieu of flowers, donations in Stark's memory can be made to the Golden Gate National Parks Conservancy, Fund in Memory of Lawrence W. Stark, Attn: Audrey Yee, Fort Mason Building 201, San Francisco, CA 94123-0022. In accordance with Stark's wishes, the money will be used to purchase land that will be kept preserved and open to the public.


Return to ME Home Page